IMAGE-BASED RECOGNITION, 3D LOCALIZATION, AND RETRO-REFLECTIVITY EVALUATION OF HIGH-QUANTITY LOW-COST ROADWAY ASSETS FOR ENHANCED CONDITION ASSESSMENT

BY

VAHID BALALI

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Civil Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Assistant Professor Mani Golparvar-Fard, Chair
Professor Khaled El-Rayes
Professor Imad Al-Qadi
Associate Professor Liang Liu
Assistant Professor Nora El-Gohary

ABSTRACT

Systematic condition assessment of high-quantity low-cost roadway assets such as traffic signs, guardrails, and pavement markings requires frequent reporting on location and up-to-date status of these assets. Today, most Departments of Transportation (DOTs) in the US collect data using camera-mounted vehicles to filter, annotate, organize, and present the data necessary for these assessments. However, the cost and complexity of the collection, analysis, and reporting as-is conditions result in sparse and infrequent monitoring. Thus, some of the gains in efficiency are consumed by monitoring costs. This dissertation proposes to improve frequency, detail, and applicability of image-based condition assessment via automating detection, classification, and 3D localization of multiple types of high-quantity low-cost roadway assets using both images collected by the DOTs and online databases such Google Street View Images. To address the new requirements of US Federal Highway Administration (FHWA), a new method is also developed that simulates nighttime visibility of traffic signs from images taken during daytime and measures their retro-reflectivity condition.

To initiate detection and classification of high-quantity low-cost roadway assets from street-level images, a number of algorithms are proposed that automatically segment and localize high-level asset categories in 3D. The first set of algorithms focus on the task of detecting and segmenting assets at high-level categories. More specifically, a method based on Semantic Texton Forest classifiers, segments each geo-registered 2D video frame at the pixel-level based on shape, texture, and color. A Structure from Motion (SfM) procedure reconstructs the road and its assets in 3D. Next, a voting scheme assigns the most observed asset category to each point in 3D. The experimental results from application of this method are promising, nevertheless because this method relies on using supervised ground-truth pixel labels for training purposes, scaling it to various types of assets is challenging. To address this issue, a non-parametric image parsing method is proposed that leverages lazy learning scheme for segmentation and recognition of roadway assets. The semi-supervised technique used in the proposed method does not need training and provides ground truth data in a more efficient manner. It is easily scalable to thousands of video frames captured during data collection. Once the high-level asset categories are detected, specific techniques needs to be exploited to detect and classify the assets at a higher level of granularity. To this end, performance of three computer vision algorithms are evaluated for

classification of traffic signs in presence of cluttered backgrounds and static and dynamic occlusions. Without making any prior assumptions about the location of traffic signs in 2D, the best performing method uses histograms of oriented gradients and color together with multiple one-vs-all Support Vector Machines, and classifies these assets into warning, regulatory, stop, and yield sign categories. To minimize the reliance on visual data collected by the DOTs and improve frequency and applicability of condition assessment, a new end-to-end procedure is presented that applies the above algorithms and creates comprehensive inventory of traffic signs using Google Street View images. By processing images extracted using Google Street View API and discriminative classification scores from all images that see a sign, the most probable 3D location of each traffic sign is derived and is shown on the Google Earth using a dynamic heat map. A data card containing information about location, type, and condition of each detected traffic sign is also created. Finally, a computer vision-based algorithm is proposed that measures retro-reflectivity of traffic signs during daytime using a vehicle mounted device. The algorithm simulates nighttime visibility of traffic signs from images taken during daytime and measures their retro-reflectivity. The technique is faster, cheaper, and safer compared to the state-of-the-art as it neither requires nighttime operation nor requires manual sign inspection. It also satisfies measurement guidelines set forth by FHWA both in terms of granularity and accuracy.

To validate the techniques, new detailed video datasets and their ground-truth were generated from 2.2-mile smart road research facility and two interstate highways in the US. The comprehensive dataset contains over 11,000 annotated U.S. traffic sign images and exhibits large variations in sign pose, scale, background, illumination, and occlusion conditions. The performance of all algorithms were examined using these datasets. For retro-reflectivity measurement of traffic signs, experiments were conducted at different times of day and for different distances. Results were compared with a method recommended by ASTM standards. The experimental results show promise in scalability of these methods to reduce the time and effort required for developing road inventories, especially for those assets such as guardrails and traffic lights that are not typically considered in 2D asset recognition methods and also multiple categories of traffic signs. The applicability of Google Street View Images for inventory management purposes and also the technique for retro-reflectivity measurement during daytime demonstrate strong potential in lowering inspection costs and improving safety in practical applications.

ACKNOWLEDGMENTS

I would like to express profound gratitude to my advisor Dr. Mani Golparvar-Fard for giving me the opportunity to be a part of this project and all the guidance and support he has given me along the way. I am inspired by the high standards that he sets for himself and those around him, his attention to detail, his dedication to the profession, and his intense commitment to his work. I have thoroughly enjoyed working with him and feel like I have learned a lot.

I would like to give special thanks to my parents, Mostafa Balali, and Parvin Sariri, as well as my brothers who taught me the value of hard work, efficiency, and self-sufficiency.

Special thanks to Dr. de la Garza for his contributions in the first phase of my research at Virginia Tech and his continuous support during my studies to this date. I greatly appreciate the guidance from, and interaction with our collaborators, in particular Armin Ashouri Rad at Virginia Tech and Amin Sadeghi at University of Illinois. They created a great environment for sharing ideas and working together toward a common goal. It was a pleasure and an honor to work with them on this project. I would also like to thank Dr. David Forsyth, Dr. Omidreza Shoghli, and Dr. Hassan Ozer for their contributions.

I would like to thank Gregg Lawrence and Kyle Armstrong at Illinois Department of Transportation for providing me the video data collected from I-57 for years 2013 and 2014. I would also like to thank the Illinois Center for Transportation for providing different types of traffic signs with different level of retro-reflectivity. The support of Illinois Department of Transportation's Bureau of Materials and Physical Research in Springfield, IL, especially Kelly Morse, is also greatly appreciated.

I appreciate the guidance and support of my Ph.D. committee, Dr. Imad Al-Qadi, Dr. Khaled El-Rayes, Dr. Liang Liu, and Dr. Nora El-Gohary.

This material is in part based upon work supported by the Institute of Critical Technology and Applied Science (ICTAS) at Virginia Tech under Grant No. (7248) for the first two years. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the Institute of Critical Technology and Applied Science (ICTAS) or other sponsors.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1. INTRODUCTION

Roadway assets are essential physical components of an infrastructure system that require preventive, restorative, or replacement work activities to preserve their functionality in an accepted level of service. Managing and maintaining infrastructure is not a new problem, nonetheless, in recent decades significant expansion in size and complexity of the infrastructure networks have posed several new engineering and management problems on how existing infrastructure can be monitored, prioritized, and maintained in a timely fashion (Golparvar-Fard et al. 2012). The fast pace of deterioration, and the limited funding available have motivated the Departments of Transportation (DOTs) to consider prioritizing roadway assets based on their existing conditions. In the meantime, the American Society of Civil Engineers (ASCE) has estimated that $170 billion in capital investment would be needed on an annual basis to improve existing conditions of the national infrastructure system (ASCE 2013). Despite the significance, there is a lack of reliable and up-to-date databases which can integrate geospatial, economic, and maintenance asset data. Such centralized databases can help DOTs better prioritize different roadway sections for maintenance and replacement planning purposes. This requires the DOTs to always keep an updated record on the condition of many types of high-quantity low-cost roadway assets such as light poles, guardrails, pavement markings, and traffic signs (Balali et al. 2015; Cheok et al. 2010).

The key elements toward development of an asset management program that is capable of producing such inventories are: 1) inexpensive and continuous data collection; and 2) methods that can further analyze the collected data for condition assessment purposes. For many of these high quantity low capital cost assets, the records of locations and most updated status are either unavailable or incomplete (Balali et al. 2013). The DOT practitioners can then leverage these assessments for maintenance and replacement planning purposes, and ultimately improve the condition of the overall transportation systems. To minimize challenges in data collection, over the past few years, the DOTs have pro-actively looked into road inventory data collection techniques. Given the significance of the problem, the Federal Highway Administration (FHWA) has recently requested identification of the gaps between currently available data collection technologies and the need for collecting comprehensive information about the nation's roadway infrastructure (FHWA 2010). Thus, over the past few years, several US DOTs have looked into Information Technologies (IT) that enable both raw and formatted asset data to be processed,

stored, and utilized in an integrated asset management system. For roadway asset management, today's most common IT capabilities focus on collecting inventory data (asset location, quality, age), typically together with photographic documentation, and also tracking public comments on current conditions (Flintsch and Bryant 2009; Markow 2007). For example, the Tennessee Department of Transportation receives continuous updates from Tennessee Road Information Management System (TRIMS) and Maintenance Management System (MMS) in a central database of roadway assets including traffic signs, guardrails, and pavement markings. The New Mexico State Highway and Transportation Department collects data on most types of visible roadway assets except for light posts and road detectors (Haas and Hensing 2005). Virginia Department of Transportation has also recently developed a web-based asset management system using Google maps and Google earth (de la Garza et al. 2010). While there is evident documented benefits that these methods address problems in data collection, the process of identifying assets and inspecting their availability, exact locations, and conditions remains dominantly manual and still needs to be systematically addressed.

Despite the importance of the level of detail and the accuracy in the data collection process, current practices are still predominantly manual, time-consuming, labor-intensive, subjective, and potentially unsafe (Balali et al. 2013; de la Garza et al. 2010; Rasdorf et al. 2009). In addition, most maintenance decision-making approaches employ a discrete representation of asset conditions. Advances in continuous condition based decision-making are of interest to the infrastructure management community, since infrastructure damage variables are typically continuous in nature. Rapid advances in automated inspection techniques are easily measuring these damage variables, and practical benefits from considering this more natural representation of condition are increasingly possible. These advances foster further research in formulating, solving, and implementing infrastructure management methods using continuous representations of important condition variables. Some research studies have already addressed the problem of automated detection, classification, and assessment of roadway assets in a discrete fashion (Mashford et al. 2007; Meegoda et al. 2006).

The significant size of the data which needs to be collected also negatively impacts the quality of the data collection and data analysis process. In addition, the subjectivity and experience of the raters have an undoubted influence on the final assessments (Balali et al. 2013; Bianchini et al. 2010). The substantial expansion in size and complexity of roadway networks, in addition to

the difficulties in data collection has made the National Academy of Engineering (NAE 2010) to identify the process of efficiently creating records of the locations and up-to-date status of the civil infrastructure as one of the Grand Engineering Challenges of the 21st century. There is a need for a credible and well-managed asset data collection and analysis that can provide useable asset inventories to DOTs for further condition assessments. This method needs to enable inexpensive and continuous data collection for high-quantity assets and provide detailed data on their conditions. Figure 1.1 shows examples of two major categories of high-quantity, low-cost roadway assets:

- Assets that can be detected from 2D images (e.g. traffic sign),
- Assets that cannot be segmented from single imagery and need 3D data for proper segmentation (e.g. guardrail)

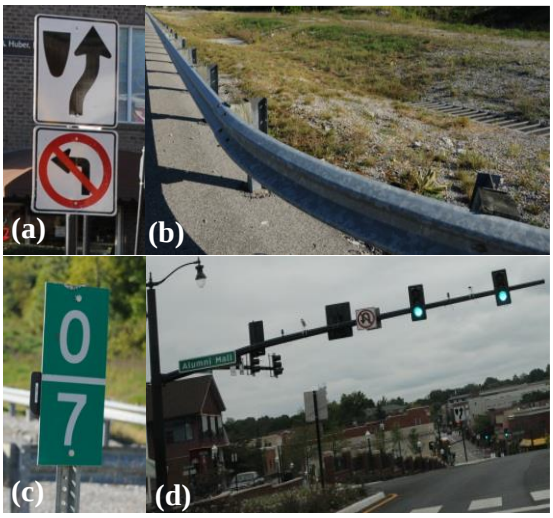

Figure 1.1 Example Frames from Video Sequences in our Asset Dataset. Assets That Can Be Detected in 2D Images: (a) Traffic Signs and (c) Mile Marker; Assets That Cannot Be Detected from Single Images and Need to Be Detected In 3D: (b) Guardrails and (d) Traffic Light

Furthermore, data collection methods for roadway inventory management are not standardized in the United States. In fact, most state DOTs still have to drive a vehicle down the road of interest and record any observed problem manually (Balali et al. 2013; Brkic 2010; de la Garza et al. 2011). To streamline the process of data collection, research has focused on application of automated identification sensors. Recent examples include application of GPS/GIS (Tsai et al. 2009), road sensors; e.g., In-Vehicle Navigation Service (INS) (Jalayer et al. 2013), Inertia Measurement Unit (IMU) (Gong et al. 2012), and Distance Measurement Indicator (DMI)

(Mordohai et al. 2007; Scaramuzza et al. 2009) , laser scanning (de la Garza et al. 2011; Walters and Jaselskis 2005), and RFID (de la Garza et al. 2010).

Today, the most dominant technique involves videotaping road assets on a massive scale using inspection vehicles equipped with three to five frontal high-resolution cameras (Balali and Golparvar-Fard 2014). These videos contain significant visual information about all assets (type, location, and condition) and their supporting infrastructure systems (poles, trusses, etc.) which make them very appealing for remote condition assessment purposes. Typically, by sitting in front of two or more monitor screens, practitioners detect and assess the condition of these assets based on their own experience and a condition assessment handbook. Due to the high cost of manual condition assessment for millions of miles of roads, assessments are only conducted for critical roadways and are performed intermittently. Thus, the records on many local and regional roads are often not updated frequently. Video-based data collection and analysis has to be done for millions of miles of roads and the practice needs to be repeated periodically.

Recently several non-DOT entities have commenced collecting street-level panoramic photographs on a country-wide scale. Examples include Google Street View, Microsoft Streetside, Mapjack, EveryScape, and Cyclomedia Globspotter (Creusen and Hazelhoff 2012). The availability of these large-scale databases– which are also frequently updated– offers the possibility to replace or perhaps augment the current DOT practices of roadway asset data collection and minimize costs (Balali et al. 2015). In particular, using Google Street View images can reduce the number of redundant enterprise information systems that collect and manage traffic inventories. Applying computer vision methods to these large collections of images has potential to create the necessary inventories more efficiently (Balali et al. 2015). One has to keep in mind that beyond changes in illumination, clutter/occlusions, varying positions and orientations, the intra-class variability can challenge the task of automated traffic sign detection and classification.

The level of detail and accuracy required for a high-quantity road asset data collection to document locations, physical attributes, and existing conditions, primarily depends on the intended use of the data (Flintsch and Bryant 2009). The process of condition assessment using massive visual datasets that the DOTs are collecting today involves reviewing all videos, manually detecting and localizing each asset in relevant video frames, corresponding them to prior assessments (if such database exists), and then performing manual condition assessment based on visual observations (FHWA 2005). In todays practice, the first few steps are the main bottlenecks

of the process. Instead of manually detecting and localizing assets within video frames and matching them to prior assessments which according to our verbal conversations with experts from Virginia and Illinois DOTs can take up to 60-70% of their time ideally the experts would only spend their time on the more value adding tasks of performing condition assessment on already detected assets, and decide on how existing conditions can be improved. Due to high costs associate with the reviews, the number of inspection cycles are very limited (MNDOT 2009) e.g. a survey cycle of one year duration for critical roadways. This creates negligence for all other local and regional roads which are also frequently used by commuters. Hence, many critical decisions may be made based on inaccurate or incomplete information, which may ultimately affect the assets maintenance and rehabilitation process.

Instead of introducing a new method for data collection, this research leverages existing and already available video frames collected by the DOTs. These videos have high qualities and in particular high spatial resolution making them ideal for computer vision method. These videos depict large number of similar assets from different camera locations and viewpoints, and have wide variability in terms of illumination conditions, and video resolution/quality. Another challenge is the intra-class variability in the visual appearance of the road assets. During the data collection, occlusions are also frequent, and asset positions and orientations may vary (Balali et al. 2013). Automating the analysis of massive visual datasets for detecting, localizing, and analyzing condition of road assets is a challenging research problem. Thus, in this research, we propose a new solutions that facilitate the processing of these existing videos. Such system has potential to minimize the need for detection and identifying asset in each video frame, and allows the expert to focus on the more important task of condition assessment.

Traffic signs exhibit large pool of inter and intra-class similarities – As shown in the US Manual on Uniform Traffic Control Devices (MUTCD) ((FHWA) 2003), traffic signs are fabricated with large variety in appearance including materials, shapes, sizes, legends, and colors (Balali and Golparvar-Fard 2015; Tsai et al. 2009a). The MUTCD contains a few hundred different signs which are divided into 13 categories. Examples of inter and intra-class variability among the US signs can be seen in Figure 1.2. Many traffic signs in the US are visually alike and unlike the European ones, they primarily rely on text as opposed to variations in color, shape, and symbol. A robust detection and recognition algorithm should detect non-standardized assets (including traffic

signs) given hundreds of variations in geometrical properties (shape, dimension), or appearance (color, text, or font).

Figure 1.2 Intra-class Variability of Traffic Signs in Different Illumination

As a first step in addressing these challenges, development of vision-based algorithms in recent years has primarily focused on detection and classification of assets in a discrete fashion. Some of these research projects – for examples those funded by the Defense Advanced Research Project Agency (DARPA 2012)– are motivated by the needs for autonomous vehicle systems. On the commercial side, several companies such as Google has also focused on the task of detection and classification of traffic signs for autonomous navigation purposes. In the case of Google autonomous vehicle, the joint application of laser scanners and cameras are proposed (Ali et al. 2014; Cimpoi 2011). Consequently, several high-end vehicles are already equipped with driver assistance systems which offer automated detection and classification for a few classes of traffic signs (Brkic 2013). However the significance of the condition assessment task and the technical challenges are different. Not only a wide range of traffic signs need to be detected, but also both False Positive (FP) and False Negative (FN) rates should be very low. Ideally a method should also leverage the already collected video streams, because of many existing contracts for video data collection, allowing for less expensive alternatives for condition assessment purposes. These are the main reasons why current traffic sign inventory and condition assessment practices are still carried out manually. Accurate and inexpensive traffic sign detection can provide more frequent condition assessment of these type of assets over the time and such database would be useful for forecasting the performance and reliability assessment.

The first step in managing these assets is monitoring their as-is conditions which involves evaluating placement, message clarity, line-of-sight, redundancy, daytime color, and nighttime

visibility (Balali et al. 2015). Nighttime visibility depends on a material property called retro-reflectivity. Due to the significance of the nighttime performance, the U.S. Federal Highway Administration (FHWA) has enacted new regulations on minimum levels of retro-reflectivity for all traffic signs. As of January 2015, all agencies are required to comply with these requirements for red/white "regulatory" signs such as Stop and Speed Limit signs, yellow "warning" signs, and green/white "guide" signs. By January 2018, these requirements will also be applicable to overhead guide signs and street name signs (Carlson and Picha 2009).

Enforcing these regulations requires agencies to frequently measure the current levels of retro-reflectivity and devise replacement plans. Two major measurement techniques are commonly used today: (a) using remote retro-reflectivity measurement device mounted on inspection vehicles. This device automatically measures retro-reflectivity, though the process needs to be done at night; (b) a practitioner using a hand-held device to measure retro-reflectivity. This device must physically touch the sign so the practitioner needs to perform this operation manually and sign-by-sign. This process can be performed during daytime, however, it is time-consuming, unsafe, and expensive (Balali and Golparvar-Fard 2015; Khalilikhah et al. 2015).

To address current limitations, we propose a new technique which is similar to (a) as it performs remote measurements and is similar to (b) as it can be used during daytime. More precisely, we use computer vision techniques to reconstruct nighttime images using images taken during the day. We then use the reconstructed night images to measure retro-reflectivity similar to (a).

Retro-reflectivity is a property of surface materials that reflect light transmitted by a distant source (Austin et al. 2009). Retro-reflective materials are commonly used for traffic signs to enhance their visibility during night. However, a perfect retro-reflector reflects all the incoming light back toward the headlights, and can cause a safety hazard for the drivers. To create optimum visibility and to divert the reflected light from the driver's direct line of sight, signs are coated with certain matt sheeting materials that have prismatic and micro-prismatic patterns. This coating mechanism allows the signs to be located safely out of the incoming light's line of travel and yet be visible to the drivers at night. Figure 1.3 illustrates the role of retro-reflective material in sign visibility.

Figure 1.3 Some Examples of Daytime (Left) and Nighttime (Right) View of Traffic Signs

According to the FHWA's Manual on Uniform Traffic Control Devices (MUTCD) (FHWA 2009), minimum retro-reflectivity requirement depends on the color combination of a sign (Table 1.1). These guidelines –current as of January 2015– require retro-reflectivity to be measured in $(cd/lx/m^2)$ at an observation angle of 0.2˚ and an entrance angle of -4.0˚. Here cd is candela– the unit of luminous intensity, which is the power that is emitted by the light source in a particular direction (brightness of a display devices) and lx is the unit of illuminance, which is a measure of how much the incident light illuminates the surface (hits and passes through a surface).

Table 1.1 Minimum Retro-reflectivity Levels from the MUTCD. Retro-reflectivity Levels Are Measured in $(cd/lx/m^2)$ at an Observation Angle of 0.2˚ and an Entrance Angle of -4.0˚

Sign Color	Sheeting Type (ASTM D4956-04)				Additional Criteria
	Beaded Sheeting			Prismatic Sheeting	
	I	II	III	III, IV, VI, VII, VIII, IX, X	
White on Green	W*; G≥7	W*; G≥15	W*; G≥25	W≥250; G≥25	Overhead
	W*; G≥7	W≥120; G≥15			Ground-mounted
Black on Yellow	Y*; O*	Y≥50; O≥50			
Black on Orange	Y*; O*	Y≥75; O≥75			
White on Red	W≥35; R≥7				
Black on White	W≥50				

The latest MUTCD recommends two approaches for retro-reflectivity management (Figure 1.4). **Management methods** which are primarily based on the life expectancy of the overall sign inventory. In these methods, life expectancy is estimated based on a number of factors including warranties, demonstrated performance, or control sign assessments. **Assessment methods** involving regular nighttime visibility measurements. The MUTCD permits the combination of these methods in any responsible program that reasonably assures compliance.

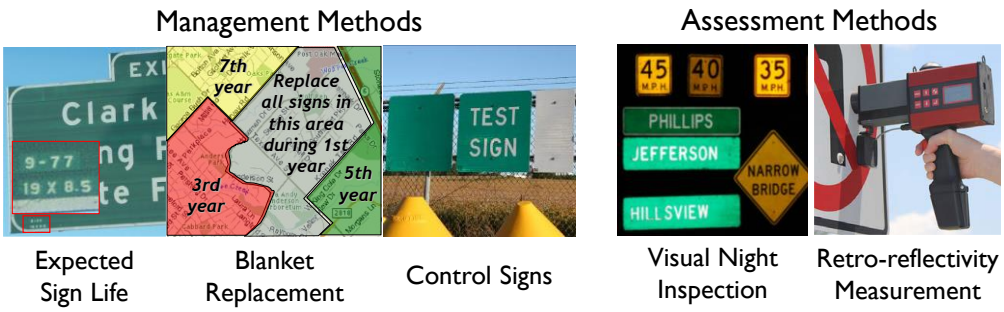

Figure 1.4 Retro-reflectivity Maintenance Methods

There are several commonly used management methods which are as follows:

Expected Sign Life Method – Using various measures of demonstrated sheeting life, signs are replaced when they reach a certain age. This method requires agencies to track the installation dates using stickers on the back of the signs, bar codes, or computerized sign management systems. However, manually inspecting these stickers or barcodes can be very time-consuming, especially if the stickers or barcodes are not easily visible on the sign.

Blanket Replacement Method is similar to expected sign life method, except that individual signs are not tracked. Instead a groups of signs are replaced at the same time based on the location and/or the type of the signs. In this method, as shown in Figure 1.4, an agency divides their jurisdiction into corridors and/or zones where the number of areas is related to replacement cycles. Then all signs are replaced in each zone/corridor according to their replacement cycles.

In **Control Signs Method**, for a groups of similar signs, a single representative sign is placed at a controlled location. The control sign is measured for retro-reflectivity periodically. When the control sign is near minimum retro-reflectivity requirement, its in-service companions are replaced. This method has a low cost and does not require labor-intensive processes. However, a large set of control signs and adequate space are required to create statically significant results.

Even though in management-based methods, signs are assigned a life expectancy at installation, yet their degradation speed varies based on location and environmental conditions. Thus, management methods which replace signs according to a schedule; will result in replacement of many signs that are still fully functional.

Compared to management methods, the assessment methods lead to less frequent sign replacements and improve efficiency. These methods could be performed through visual inspection during nighttime (**Nighttime Visual Inspection Method**) or using retro-reflectometer

devices during daytime (**Retro-reflectivity Measurement Method**). Through these methods, retro-reflectivity could be measured more regularly, and recommendations can be provided on the best timing for replacing a sign based on degradation levels. Manual visual inspections involve some degree of subjectivity, yet research has shown that trained observers can reasonably identify signs with marginal retro-reflectivity (Carlson and Lupes 2007). Measuring sign retro-reflectivity through a systematic process provides the most direct means of monitoring the maintained retro-reflectivity levels and removes subjectivity.

Currently, the most objective method to measure retro-reflectivity is to use handheld retro-reflectometers (Brimley and Ye 2013; Hummer et al. 2013). A single handheld retro-reflectometer costs over ten thousand dollars (Reynolds 2012). Because this device needs to be in direct contact with a sign surface to take measurements - even for the overhead signs and those ground mounted signs that are out of reach (see Figure 1.5)- its application can be labor-intensive, expensive, and potentially unsafe for the inspectors (Balali and Golparvar-Fard 2014; Preston et al. 2014). A bigger challenge is that these devices can only take measurements on pre-defined geometries which are not the best representatives of the actual driving geometries (Babić et al. 2014; Bhalla et al. 2003; Hulme et al. 2011). For example, measurements from twisted and leaning signs can result in retro-reflectivity above the minimum levels, while the actual luminance of the sign under nighttime conditions may be lower than the requirements (Shcukanec et al. 2014). Overall, all of the management or assessment methods, even those that are standardized by (FHWA 2009), are costly and/or labor intensive. In the past decade, research has proposed several measurement techniques to address these shortcomings. These techniques are covered in Chapter 2.

Figure 1.5 Different Types of Handheld Retro-reflectometers. The Operator Needs to Assess Retro-reflectivity of Each Sign Separately

1.1. Research Objectives

The overarching goal of this research is to automate the entirety of data analysis and management of the low-cost high-quantity roadway assets specifically focusing on traffic signs by exploiting the application of a mounted array of inexpensive cameras and computing capacity on roadway inspection vehicles. Given the proposed configuration of cameras, the scope of the proposed research includes assets that are mainly located in the front and the right side of the mounted array of cameras: traffic signs, mile markers, pavement markings, traffic signals, light poles, guardrails and guardrail terminals. As a result detection of pavement distress, drop-inlets, paved ditches, and unpaved shoulders that are primarily visible on a road surface is not part of the scope of this research. The proposed framework can provide asset management and condition assessment researchers and evaluators with an accurate and comprehensive database of all types of traffic signs, allowing the former to build on this research towards the automation of condition assessment, and the latter to make better informed decisions on condition assessment and the best timing and strategies for maintenance. The goal of this research will be accomplished through the following research objectives:

- **Objective 1:** Segmentation and 3D reconstruction of roadway assets
    - Create and validate a method for creating a 3D point cloud model of all objects along the roadway with the same video frames
    - Develop a method for identification of potential areas in 3D for extraction of particular types of assets
    - Refine the classification of assets and their types using joint appearance and 3D shape detection with high accuracy
    - Develop a method for localizing detected assets in the reconstructed 3D point cloud models
- **Objective 2:** Segmentation and recognition of roadway assets from car-mounted camera video streams
- **Objective 3:** Evaluation of multi-class traffic signs detection and classification
    - Create and validate a method for identifying potential areas in 2D images for asset candidate extraction
    - Create and validate a method for classifying the 2D shape of the asset candidates with reasonable accuracy
    - Create and validate a method for classifying the texture and color of the asset candidates with reasonable accuracy
- **Objective 4:** Mapping and 3D localizing of traffic signs using Google Street View images

- Develop a method for visualizing the assets and their types in an augmented reality environment for researchers and practitioners

- **Objective 5:** Image-based retro-reflectivity measurement of traffic signs in daytime

Develop a method for measuring the retro-reflectivity of traffic sign in a daytime and remotely

1.2. Dissertation Structure

This dissertation describes some of the segmentation, detection, classification, and development of the image-based 3D reconstruction and recognition of high-quantity low-cost roadway assets for enhancing the condition assessment. It is organized as follows:

- Chapter 1 provides background including introduction and current practices in roadway asset management and why photogrammetry technology is needed for recognition, reconstruction, and condition assessment.
- Chapter 2 presents a review of previous research into the development of image-based recognition and 3D reconstruction of roadways for enhanced condition assessment.
- Chapter 3 presents the detail of the propose methods for segmentation, 3D reconstruction, detection, classification, localization, and condition assessment.
- Chapter 4 shows the results and presents detailed discussion on each of the new method introduced in this dissertation
- Chapter 5 summarizes the applicability of the proposed methods and finishes the conclusions and future direction from this research.

CHAPTER 2. LITERATURE REVIEW

In most state-of-the-art practices, the analysis of the roadway assets data is not fully automated. The significant amount of information required to be manually processed may adversely affect the quality of the analysis, resulting in subjective reports (Torrent and Caldas 2009), and minimizes opportunities for continuous monitoring which is a necessary step for roadway asset management improvement. Hence, many critical decisions may be made based on inaccurate or incomplete information, ultimately affecting the assets' maintenance and rehabilitation process.

Data collection, data management, and data integration are essential parts of a successful roadway asset management program (Flintsch and Bryant 2009). Important issues in the data collection and analysis process include accuracy and cost, level of subjectivity and variability, the speed of the process, as well as safety of the inspection crew.

2.1. Data Collection

Most of the roadway inventory data collection methods use one or more visual sensing methods to capture road inventory information. GPS, Inertia Measurement Unit (IMU), and Distance Measurement Indicator (DMI) are often used to provide accurate positional data for these visual sensing systems (Balali et al. 2013; Gong et al. 2012). The level of detail and accuracy of data collection primarily depend on the intended use of the data, yet in almost all cases can be classified into three categories (Flintsch and Bryant 2009):

- Location of the asset;
- Physical attributes: e.g., description of the asset, material type, size, length;
- Condition: qualitative and generic: e.g., good or bad; quantitative and detailed: e.g., asset condition index.

There are currently four dominant practices for asset data collection which are as follows:

2.1.1. Manual data collection

Automated methods that can facilitate the entire process of roadway asset data collection, verification, and updating are extremely limited or non-existent (de la Garza et al. 2011; FHWA 2010). A large number of local and state highway agencies rely on extensive use of inspection

crews for manpowered data collection. As a result, the process of collecting roadway and asset data is conducted on a sporadic basis (FHWA 2010). With an exhaustive manual data collection, every necessary detail can be captured; nonetheless, the process will be extremely labor-intensive, time-consuming, costly, and potentially unsafe (Tsai and Wang 2008; Uslu et al. 2011). This in turn reduces the chances of frequent updates on the condition of assets. There is a need for a cost-effective method that can collect such data automatically.

2.1.2. Semi-automated data collection

Semi-automated methods could facilitate data collection, but still require manual post-processing for the analysis of the collected data. This further limits their application for assets that are located across thousands of road miles. Examples of these technologies include handheld computers equipped with GPS, road sensors, barcodes, and RFID tags. The collected data are documented either with pen and paper, or in more recent cases, with handheld computers equipped with GPS (de la Garza et al. 2009; Larson and Skrypczuk 2004). Despite the benefits of an electronic database, data logging is still conducted manually. In 2009, VDOT also conducted a pilot project for photographic documentation for all asset failures within the Stanton South TAMS Project (de la Garza et al. 2010). In both of these cases, the data collection process is almost fully automated, yet manual reviewing of millions of video frames to extract roadway assets is considerably labor-intensive and costly.

2.1.3. Automated data collection

The availability of cheap and high-resolution video cameras, large data storage capacities, in addition to advances in computing has resulted in most state DOTs adopting photographic documentation for their roadway assets (Hu and Tsai 2011b; Rasdorf et al. 2009). For example, the New Mexico State Highway and Transportation Department (NMSHTD) collects data on most types of visible roadway assets except for light posts and road detectors. As a result, a comprehensive report of certain types of assets can be provided on a regular basis (FHWA 2010). The data is collected using a sophisticated digital inspection vehicle which is equipped with four cameras mounted on support bars on the roof of the van. Their Road Feature Inventory (RFI) is then visualized by a "Virtual Drive" method. A user can conduct an inspection walk through of

the route in a virtual environment and analyze the types, locations, and conditions of the assets manually.

2.1.4. Remote data collection

The last method pertains to the use of satellite imagery and remote sensing application. With advances in remote sensing technologies, it is possible to collect different types of data pertaining to the surface of Earth with relative ease. The Center for Transportation Research and Education (CTRE) is investigating the potential of high-resolution images acquired through satellites, lasers, and aerial photos for various aspect of roadway asset management (CTRE 2004). The images are used in conjunction with ground information in order to reference the location of assets and assess their conditions (NASA 2000; NCRST 2001). Similar to the previous data collection techniques, analyzing and evaluating conditions of the roadway assets is still predominantly manual and involves time consuming processes.

2.1.5. Existing technologies in data collection

Nowadays, various automated identification technologies can be utilized to enhance data collection, identification, and tracking of components in infrastructure management such as GPS and GIS (Tsai et al. 2009a), road sensors (Scaramuzza et al. 2009), laser scanning (de la Garza et al. 2011; Jaselskis et al. 2006), Radio-Frequency IDentification (RFID) (de la Garza et al. 2009; Kiziltas et al. 2008), barcode, Optical Character Recognition (OCR), and contact memory technologies. Other recent advancements in sensor and computer technology such as lasers, visual and infrared cameras, and ultrasound have also created new opportunities to collect data and improve the roadway infrastructure system more efficiently and accurately. Computer vision and image based reconstruction are particularly the latest techniques that have been used for automated detection, classification and assessment of assets in a discrete fashion. Some practices focus on condition assessment of assets. Examples of these works include (Hu and Tsai 2011; Tsai et al. 2009b) which focus on condition assessment of traffic signs, or (Meegoda et al. 2006) which is an algorithm for condition assessment of culverts.

2.1.6. Vision-based technologies

a. Laser scanners

Advances in sensor technology, such as lasers have enabled capturing surface and subsurface roadway conditions and collect data at much higher speed. A laser scanner is a 3D imaging technology designed for capturing vast amounts of measurements of points in its vicinity in a short period of time. 3D laser scanning technology is being increasingly used for constructing 3D virtual representations of infrastructures. A 3D laser scanner calculates the distance between the object and the scanner by emitting a laser beam either by determining the phase difference between the emitted and returned signals (phase-based scanners) or by calculating the laser beam travel time (time-of-flight scanners). The most common type of 3D imaging systems currently used in construction is time-of-flight laser scanners (Kavulya et al. 2011; Kiziltas et al. 2008). Phase-based laser scanners theoretically achieve measurements with higher accuracy as compared to time-of-flight scanners.

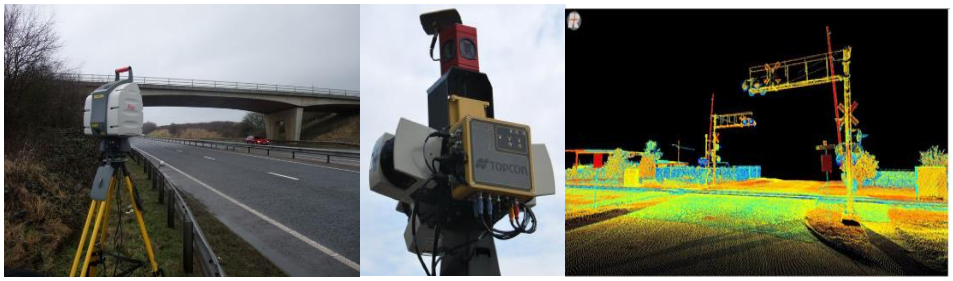

Figure 2.1 Laser Scanners and Generated Point Cloud Models

In a recent study, (de la Garza et al. 2011) implemented an Integrated Positioning System for 3D Mobile Mapping (IP-S2) to rapidly capture high-resolution 3D data of all roadway assets. Such systems enable data to be collected in a short period of time, and the vehicle speed has little impact on the quality of data. Nonetheless, the identification of the assets from the 3D data still needs to be conducted manually. Different lighting conditions and distances from the vehicle can also adversely affect the quality of the data. The methods that use laser scanning present other limitations such as high initial cost requirement, need for significant power, frequent maintenance, and expertise for operating the device. Laser scanners also suffer from mixed-pixel phenomena and require noise to be manually removed from the data in a post processed stage. Finally, laser scanners only provide Cartesian point clouds and do not enable other semantics such as appearance which is necessary for asset classification to be easily extracted. The laser scanner technology

cannot assess certain failure codes of high-quantity assets, including missing guardrail bolts, damage to the back of guardrail components, turned signs, and missing object markers.

b. Image/Video streams

Image-based 3D reconstruction and photogrammetric techniques enable extraction of semantics through registered imagery and as a result create unique opportunities for asset detection and localization (Balali et al. 2015; Balali et al. 2015; Golparvar-Fard et al. 2012). (Wu and Tsai 2006b) used real digital images for automated road geometric data collection, especially in recognizing lane marks and shoulder edges. In the context of infrastructure 3D reconstruction, (Brilakis et al. 2011) proposed a structured video-grammetry for 3D reconstruction of infrastructure. Most of the work in this area only focuses on 3D reconstruction for generating maps and 3D terrain models. Nonetheless, none of these vision-based methods has been used to recognize, locate, assess condition of the roadway assets, and ultimately visualize the most updated status in a 3D environment. In the past few years, a few research groups have started using 3D image-based reconstruction algorithms for identification and localization of roadway assets. (Balali et al. 2015; Timofte et al. 2014) proposed a new approach for 2D recognition and 3D localization of traffic signs. In a more recent work, (Golparvar-Fard et al. 2012) builds 3D reconstruction and validates the applicability of this algorithm for roadway assets. A dense point cloud of assets creates an opportunity for 3D segmentation of the point cloud and identification of 3D continuous assets.

c. Multi-sensors and data fusion

To address the need to collect roadway information economically, accurately, and reliably, FHWA Office of Advanced Research initiated the development of the Digital Highway Measurement System (DHMS) in 2003. DHMS is an instrumented vehicle which combined laser scanning sensors, high accuracy nationwide differential GPS and an airline quality inertial navigation unit, in order to accurately measure roadway geometry and to build 3D maps of features of interest on, over, or beside the road.

Effective management of roadway assets requires comprehensive data on the asset inventory, its current condition, and its historical performance. New Mexico State Highway and Transportation Department (NMSHTD) checks the condition of a portion of the state roadway in

remote regions using virtual drive feature of the agency's road features inventory. They use an image-based GPS data collection system for most types of visible roadway assets except for light posts and road detectors (Figure 2.2). This vehicle is equipped with four cameras mounted on support bars on the roof of the inspection vehicle. This van also carries other equipment such as a laser scanner which is used to capture pavement conditions and road geometry. By choosing a route, and start and end mileposts, a user can conduct an inspection walk through of the route on computer screen and analyze the types, location, and condition of assets manually (Medina et al. 2009).

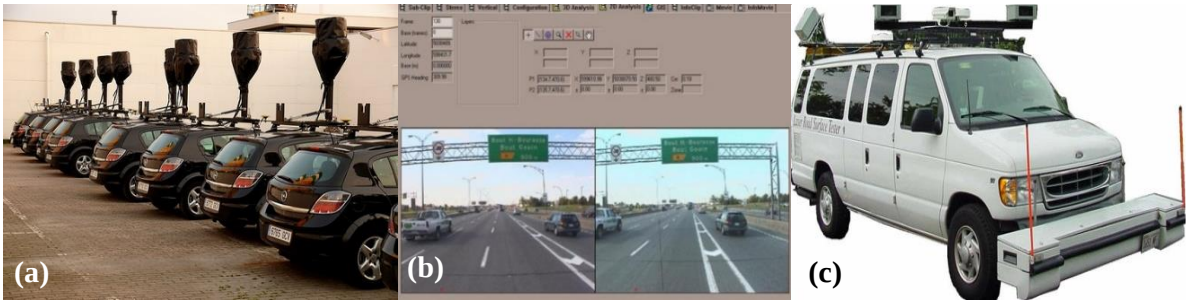

Figure 2.2 (a) Digital Highway Measurement System; (b) NMSHTD Virtual Reality; (c) NMSHTD Van

To date, state DOTs and local agencies in the U.S. have used a variety of roadway inventory methods. These methods vary based on cost, equipment type, the time requirements for data collection and data reduction, and can be categorized into four categories of field inventory methods, photo/video logs, integrated GPS/GIS mapping system, and aerial/satellite photography as shown in Table 2.1.

A nationwide survey was recently conducted by the California Department of Transportation to investigate popularity of these methods among practitioners. The results (Ravani et al. 2009) shows the integrated GPS/GIS mapping method is considered to be the best short-term solution. Nevertheless, remote sensing methods such as satellite imagery and photo/video logs were indicated as the most attractive long-term solutions. The report also emphasizes that there is no one-size-fits-all approach for asset data collection. Rather the most appropriate approach depends on an agency's needs and culture as well as the availability of economic, technological, and human resources. (de la Garza et al. 2010; Haas and Hensing 2005; Jalayer et al. 2013) have shown that the utility of a particular inventory technique depends on the type of features to be collected such as location, sign type, spatial measurement, and material property visual

measurement. As shown in Table 2.2, in all these cases the data is still collected and analyzed manually and thus inventory databases cannot be quickly or frequently updated.

Table 2.1 Existing Roadway Inventory Data Collection Methods and Related Studies

Methods	Description	Related works
Field Inventory	Using GPS survey and conventional optical equipment to collect desired information in the field	(Jones 2004; Khattak et al. 2000; Zhou et al. 2013)
Photo/Video Log	Driving a vehicle along the roadway while automatically recording photos/videos, which can be examined later to extract information	(Ai and Tsai 2014; Ai and Tsai 2011; DeGray and Hancock 2002; Hu et al. 2004; Jeyapalan 2004; Jeyapalan and Jaselskis 2002; Maerz and McKenna 1999; Robyak and Orvets 2004; Tsai et al. 2009a; Wang et al. 2010; Wu and Tsai 2006a)
Integrated GPS/GIS Mapping Systems	Using an integrated GPS/GIS field data logger to record and store inventory information	(Caddell et al. 2009; Jones 2004)
Aerial/Satellite Photography	Analyzing high resolution images taken from aircraft or satellites to identify and extract roadway inventory information	(Veneziano et al. 2002)

A nationwide survey was recently conducted by the California Department of Transportation to investigate popularity of these methods among practitioners. The results (Ravani et al. 2009) shows the integrated GPS/GIS mapping method is considered to be the best short-term solution. Nevertheless, remote sensing methods such as satellite imagery and photo/video logs were indicated as the most attractive long-term solutions. The report also emphasizes that there is no one-size-fits-all approach for asset data collection. Rather the most appropriate approach depends on an agency's needs and culture as well as the availability of economic, technological, and human resources. (de la Garza et al. 2010; Haas and Hensing 2005; Jalayer et al. 2013) have shown that the utility of a particular inventory technique depends on the type of features to be collected such as location, sign type, spatial measurement, and material property visual measurement. As shown in Table 2.2, in all these cases the data is still collected and analyzed manually and thus inventory databases cannot be quickly or frequently updated.

Table 2.2 Examples of State DOT Road Inventory Programs

State DOT	Inventory Techniques		Inventory Data
	Collection	Storage	
Washington	Photo log, integrated GPS/GIS mapping systems	GIS	Cable barriers, concrete barriers, culverts, culvert ends, ditches, drainage inlets, glare screens, guardrails, impact attenuators, miscellaneous fixed objects, pipe ends, pedestals, roadside slope, rock outcroppings, special-use barriers, supports, trees, tree groupings, walls
Michigan	Integrated GPS/GIS mapping systems, field inventory	GIS	Guardrails, pipes, culverts, culvert ends, catch basins, impact attenuators
Ohio	Photo log, integrated GPS/GIS mapping Systems	GIS	Wetland delineation, vegetation classification
Iowa	Airborne LiDAR, aerial photography	GIS	Landscape, sloped areas, individual counts of trees, side slope, grade, contour
Idaho	Video log	MS Access	Guardrails
Tennessee	Tennessee Road Information Management System (TRIMS), Maintenance Management System (MMS)	Central Database	Traffic signs, guardrails, and pavement markings which are manually collected.
New Mexico	Photo, Laser Scanner, and Virtual Reality System	Video	Most types of visible roadway assets except for light posts and road detectors
Virginia	Web-based asset management system using Google Maps	Google Maps	Cross pipes, ditches
FHWA Baltimore-Washington Parkway	Mobile mapping	Point Cloud Software, GIS	Corridors, signs

2.2. Data Management

Traditionally, the only way of checking the condition of roadway assets was to go out to the field. If the assets were numerous or far apart, this process would be very time consuming. Moreover, it was often difficult to locate a specific asset item failure in a given segment and finding failure was impossible since the condition of some assets can change in a short span of time (de la

Garza et al. 2010). Many state DOTs use some form of the random sampling method, which varies from one state to another. In the past few years, several Data Management systems have been developed that can facilitate the process. For example, Tennessee Department of Transportation (TDOT) benefits from Tennessee Road Information Management System (TRIMS) and Maintenance Management System (MMS) to continuously update data in a central data which is manually collected conventionally from traditional sources. MMS automatically transfers information to and from internal databases, including TRIMS database on roadway assets including roadway signs, guardrails and barriers, and pavement markings and treatments. Signals, lighting, and loop detectors are maintained at the local level (FHWA 2010). VDOT has also recently developed a web-based asset management system. This comprehensive system displays the failures of asset in their actual locations using Google maps and Google earth (de la Garza et al. 2010). Such a tool enables VDOT to check the status of any failed asset from any computer with an Internet connection. In these cases, the data still needs to be manually collected and as a result, these databases cannot be quickly updated. It is also very important to provide user-friendly asset management systems; otherwise, government roadway departments will not implement those (Mizusawa and McNeil 2006).

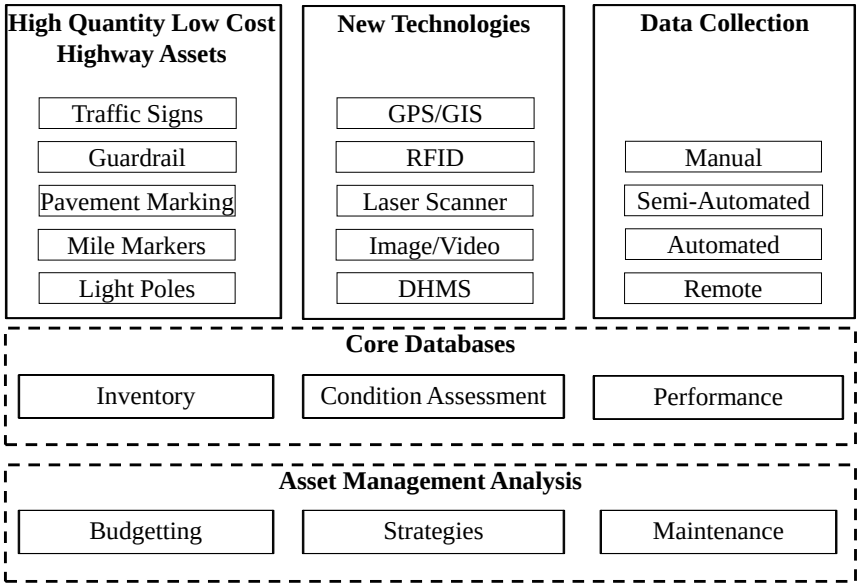

Figure 2.3 Asset Management System

2.3. Data Analysis

Despite the significance of the visual information that is embedded in video frames, to date, the full application for automated and simultaneous data collection and analysis for a wide-range

of existing assets is still unexploited by researchers. A computer vision detection method capable for a wide-range of assets must be able to segment each video frame into groups of assets, and then accurately classify and localize the detections into relevant asset categories. However, majority of the state-of-the-art methods are primarily devised to detect one type of asset. Over the past few years, research in automated data analysis has primarily focused on two aspects:

- Reconstructing 3D models of assets from images and video streams collected from cameras mounted on inspection vehicles or by using laser scanners;
- Automated recognition and classification of assets

2.3.1. Image-based / video-based 3D reconstruction

Image-based 3D reconstruction and photogrammetric techniques enable extraction of semantics through registered imagery and as a result create a unique opportunities for asset detection and localization. The state-of-the-art in image-based and video-based 3D reconstruction from images and video streams in computer vision and Architecture/Engineering/Construction and Facility Management (AEC/FM) communities have experienced several breakthroughs in the last few years. Some of more recent works in computer vision (Crandall et al. 2011; Frahm et al. 2010; Furukawa et al. 2010; Gallup et al. 2010; Heng et al. 2011; Snavely et al. 2008; Tuite et al. 2011) focus of image-based 3D reconstruction at large-scale imagery. Others such as (Mordohai et al. 2007) proposed real-time collection of videos and GPS data to produce 3D models of urban environments. (Gallup et al. 2010) presented a new multi-view depth-map fusion algorithm which attempts to produce 3D surfaces from ground-level or aerial imagery.

In the AEC/FM community, (Golparvar-Fard et al. 2009) is one of the earliest works that proposes a new Structure-from-Motion (SfM) algorithm for 3D reconstruction using unordered photo collections. More recent works (Golparvar-Fard et al. 2010; Golparvar-Fard et al. 2012b) proposed a new dense 3D reconstruction algorithm based on multi-view stereo and voxel coloring/labeling techniques which significantly improve the quality of the reconstructed models. In the context of infrastructure projects, (Uslu et al. 2011) extended the work of (Golparvar-Fard et al. 2010) and applied the method to the reconstruction of roadways and their high-quantity assets. (Brilakis et al. 2011) also proposed a structured video-grammetry for 3D reconstruction of existing roadway bridges. This technique benefits from video streams for a more complete

modeling results in sparse point cloud models of roadway assets. The low resolution of SfM point clouds may not be suitable for asset detection, localization, and condition asset management.

In the past few years, a few research groups have started using image-based 3D reconstruction algorithms for identification and localization of roadway assets. (Balali et al. 2015; Timofte et al. 2014) proposed a new approach for 2D recognition and 3D localization of traffic signs. The method primarily focuses on 3D sparse point cloud reconstruction and recognition of traffic signs, and high average performance is reported on the 2D recognition module. Yet, since it primarily focuses on 2D recognition, it is not directly applicable for 3D segmentation and classification of other types of assets including guardrails and light poles. Despite the great performance reported, the recent algorithms mainly result in sparse 3D point cloud models and as a result may not be useful for 3D detection and classification of guardrails and light poles. (Golparvar-Fard et al. 2012) proposed a new asset detection and recognition algorithm based on Semantic Texton Forest that can simultaneously segment an image and categorize assets.

2.3.2. Segmentation and recognition of roadway assets

In computer vision, image segmentation is the process of partitioning an image into multiple salient image regions in which each region can correspond to individual surfaces, objects, or natural parts of objects. To form distinct regions, these methods have conventionally focused on labeling individual pixels with an object/surface category. (Shotton et al. 2008) is among the most dominantly used methods. Their work proposes a segmentation method based on bag of semantic textons to group decision trees that can act directly on image pixels. Both textons and priors as features are used to give coherent semantic segmentation and label each pixel. The main drawback is that training generative and discriminative learning models in semantic texton forest method and other segmentation algorithms which operate at the pixel level (Ladick et al. 2010; Shotton et al. 2008; Xuming et al. 2004) that these methods are fully supervised. This requires providing a fully labeled ground-truth dataset for training purposes. The process of training can take days and must be repeated if new asset categories are added to the dataset. Processing a test image is also quite slow as it involves steps on detecting candidates over an image, performing graphical model inference, or searching over multiple segmentations.

While it is tempting to recognize objects from images, motion compensation and motion segmentation are addressed by (Golparvar-Fard et al. 2012). They proposed a new asset detection

and recognition algorithm based on Semantic Texton Forest that can simultaneously segment an image and categorize assets. The training process for the generative and discriminative learning models in the proposed STF method and many other segmentation algorithms that operate at the pixel level (Ladick et al. 2010; Shotton et al. 2008; Xuming et al. 2004) need to be fully supervised. Training can take days and must be repeated if new categories are added to the dataset. Processing a test image is also quite slow as it involves steps on detecting candidates over an image, performing graphical model inference, or searching over multiple segmentations. However, this research did not result in high accuracy rates needed for several categories of assets. More testing is yet to be conducted with comprehensive datasets.

Over the past few years, researches have focused on nonparametric and data-driven approaches that do not require significant training (Liu et al. 2011; Tighe and Lazebnik 2013). For each new test image, these methods retrieve the most similar training images and transfer the desired information from the training images to the query image for labeling. (Liu et al. 2011) proposed a non-parametric label transfer method based on estimating a dense deformation field between images using Scale Invariant Feature Transform (SIFT) flows. SIFT is an algorithm that detects and describes feature points of an image. SIFT is a robust detection and description technique which can handle changes in viewpoint, illuminations (day vs. night), and is fast and efficient enough to run in real-time. The main challenge here is the complex and expensive optimization problem associated with finding the SIFT flow. Moreover, the formulation of scene matching in terms of estimating a dense per-pixel flow field is not necessarily in accord with the intuitive understanding of scenes as collection of discrete objects based on spatial support and asset category. To address such limitation, (Tighe and Lazebnik 2013) recently proposed a non-parametric solution to image parsing that is straightforward and efficient. Their proposed method relies only on operations that can easily scale to very large collections of images and sets of labels. The more fundamental question of whether motion and 3D structure can be used to accurately segment video frames and recognize the object categories is addressed by (Brostow et al. 2008). Existing video parsing approaches (Brostow et al. 2008; Zhang et al. 2010) use structure-from-motion techniques to obtain either sparse point clouds or dense depth maps, and extract geometry-based features that can be combined with appearance-based features or used on their own to achieve greater accuracy. Their work investigated how semantic segmentation based on 3D point cloud can be derived from ego-motion information of a camera. (Tighe and Lazebnik 2013) took

a simpler approach and only used motion cues to segment the video into temporally consistent regions or super-voxels (Grundmann et al. 2010). This helps to better separate moving objects from one another especially when there is no high contrast edge between them. While it is tempting to recognize objects from images, motion compensation and motion segmentation are still open research problems, which are the basis for the work presented in this research.

The main consensus in the segmentation roadway assets is that image parsing should leverage context information (Galleguillos and Belongie 2010; Tighe and Lazebnik 2010). However, learning and inference with most current algorithms are slow. In (Golparvar-Fard et al. 2012), we explored how the challenges associated with training and testing processes for segmentation of video streams into roadway asset categories can be minimized. To devise a scalable method and minimize the need for pixel-level training, we focus on efficient forms of context that do not need training and that can be followed by super-pixel matching and efficient Makov Random Field (MRF) framework amenable to optimization for incorporating neighborhood context by fast graph cut algorithms.

2.3.3. Traffic sign detection and classification

The computer vision community has largely turned towards the recognition of object classes, rather than specific roadway assets such as traffic signs. Current research efforts in devising a computer vision model for roadway asset detection are roughly divided into three stages:

- Segmentation,
- Detection, and
- Condition assessment.

Detection of traffic signs as classified in Manual on Uniform Traffic Control Devices (MUTCD) is an area that has received considerable attention over the past few years. Traffic signs come in hundreds of variations, such as in dimension, color, text, and font. (Maldonado Bascón et al. 2010) presented a Support Vector Machine (SVM) to recognize road-signs. (Krishnan 2009) has presented a triangulation and bundle adjustment approach for identifying road signs. (Hu and Tsai 2011; Wu and Tsai 2006b) have created a nearest-neighbor assignment of feature descriptors for an image recognition model for developing a sign inventory. Although most of these techniques have achieved the goal of automation and accuracy to a reasonable level, nonetheless none of these

systems use the same visual information to locate the assets and more importantly detect them in a continuous fashion.

As a first step towards addressing this problem, research in intelligent driver assistance systems community has focused on detecting speed limit signs (Mogelmose et al. 2012). The performance of the proposed algorithms also widely varies. An earlier example is (Loy and Barnes 2004) where all signs in the testing dataset were detected successfully, however a large number of FPs rate per frame remained an open research problem. For roadway asset condition assessment, a method that can detect several different types of traffic signs at a low false detection rate is more appealing than a method that can only detect one specific traffic sign, but does that well. Table 2.3 categorizes some of the major state-of-the-art detection and classification methods based on the type of their visual features (i.e. color vs. shape).

Table 2.3 State-of-the-art Methods for Detection and Classification of Single-category Traffic Signs Categorized Based on the Type of Features

	Features	Examples from the Literature
Different Type of Features Used	Color	(Lopez and Fuentes 2007; Maldonado-Bascon et al. 2007)
	Shape	(Gil-Jim et al. 2005; Kim et al. 2005)
	Color, and Shape	(Fang et al. 2003; Gao et al. 2003; Miura et al. 2000; Shuang-dong et al. 2005)
	Geometrical, Physical Features, and Text	(Yangxing et al. 2006)
	Dimension, Color, Text, and Font	(Fatmehsan et al. 2010; Hu and Tsai 2011)
The Specific Type of Traffic Sign used for Detection	Rectangle and Triangle Shape	(Ballerini et al. 2005; Ruta et al. 2010; Shuang-dong et al. 2005)
	Stop and/or Speed Limit Signs	(Fatmehsan et al. 2010; Meuter et al. 2008; Tsai and Wu 2002; Wu and Tsai 2005; Wu and Tsai 2006b; Yea-Shuan and Yun-Shin 2010; Yea-Shuan et al. 2012)

The most recent methods in (Baro et al. 2009; Overett et al. 2011; Timofte et al. 2014) have validated their performance with reported detection rates above 90% with relatively low number of FPs. However, all these methods have been validated on European datasets and for only a few types of traffic signs. Table 2.4 summarizes the features and detection methods for these methods.

Table 2.4 Overview of the Performance for the Best Detection Rates

Paper	Features	Detection Method	Best Detection Rate	FPs for Best Detection Rate	Average Detection Rate	Average FPs	Type of Traffic Sign
(Baro et al. 2009)	Dissociated dipoles*	Cascade Classifier	97%	5.6%	92%	4.8%	Circular speed, Triangular
(Overett et al. 2011)	Histogram of Oriented Gradients (HOG)	5 Stage cascade classifier trained with LogitBoost	98.68%	10%	-	-	Circular Red signs
(Timofte et al. 2014)	Adaptive RGB threshold + Edges	Fuzzy template of a Hough derivative	95.7%	2.5%	95.29%	10.41%	Circular red and blue, Diamond white

* A more general type of features than the Haar-like features

In addition to detection, 3D localization of traffic signs from video streams has also been the focus of some of the recent works. Examples include (Soheilian et al. 2013; Timofte et al. 2014) which mainly visualize the detected signs within sparse 3D point cloud models. (Balali and Golparvar-Fard 2014; Golparvar-Fard et al. 2012) have also proposed two methods for segmentation of roadway assets at a higher-level (e.g., guardrail, signs, safety cones, etc.) based on scalable non-parametric parsing and Semantic Texton Forest algorithms, respectively. These methods can segment a video frame into different asset categories, and can serve as a basis for the task of detection and classification.

The prior work in detection and classification of traffic signs can be roughly divided into three categories of work on segmentation, feature extraction, and detection. In the following the state-of-the-art in each category is presented:

a. Segmentation and candidate extraction

The purpose of segmentation is to narrow down the search space in finding candidates for signs from the entirety of a video frame to small number of image patches (Golparvar-Fard et al. 2012). Because traffic signs have distinct colors, majority of the earlier segmentation methods focused on thresholding color channels. Since Red-Green-Blue (RGB) color space is generally

perceived to be not subject to wider variations in brightness, methods such as (FeiXiang et al. 2009; Hsin-Han et al. 2010; Wen-Jia and Chien-Chung 2007; Xu et al. 2010) leveraged the Hue-Saturation-Value (HSV) color space. Interestingly (Gomez-Moreno et al. 2010) reports that HSV-based color segmentation methods does not necessarily have a better performance against the normalized RGB color channel. In an attempt to minimize the impact of the instabilities caused by the lighting variations, (Balali et al. 2013; Prisacariu et al. 2010; Timofte et al. 2009; Timofte et al. 2014) proposed adaptive thresholds to be used on the RGB color space. While those segmentation methods that use color information perform much better than the shape-only methods, they struggle in detecting traffic signs with white background. For a more detailed comparison of the existing methods, readers are encouraged to look into (Geronimo et al. 2010).

Object detection and classification problem is traditionally solved by either the selective extraction of windows of interest, or exhaustive sliding window based classification. In the first approach small number of interest regions are selected in the images through fast and inexpensive methods. These interest regions are then subjected to a more sophisticated classification. Such approach risks overlooking some traffic signs. Second approach considers all candidate windows in the image. Given the large number of candidates, classification easily becomes intractable (Balali et al. 2013).

b. Feature extraction and detection

Because traffic signs have distinct shapes, the most dominant type of features used to-date are edges, intensity gradients, and more recently principled presentations such as Histogram of Oriented Gradients (HOG) (Alefs et al. 2007; Gao et al. 2006; Houben 2011; Mathias et al. 2013; Overett et al. 2011; Pettersson et al. 2008; Xie et al. 2009) and Haar-like features (Bahlmann et al. 2005; Baro et al. 2009; Keller et al. 2008; Prisacariu et al. 2010). (Creusen et al. 2010) also augmented the HOG feature vectors with CIELab and YCbCr color information for detecting blue-circular, red-circular and triangular signs using a relatively small training dataset (~ tens of samples per category).

c. Classification

The selection of classification method is constrained to the choice of features. The dominant methods are the Hough transform and its derivatives for model fitting (especially when

edges and intensity gradients are used). For HOG and Haar-like features, SVM, neural networks, and cascaded classifiers have been frequently reported. Particularly cascade classifiers are used more often with the Haar-like features (Bahlmann et al. 2005; Baro et al. 2009; Keller et al. 2008; Prisacariu et al. 2010). The application of HOG features with standard SVM (Creusen et al. 2010; Xie et al. 2009) and cascade classifiers with boosting variants (Overett et al. 2011; Pettersson et al. 2008) are also reported in the literature. A method using color and Haar-like features and AdaBoost cascade classifiers was also presented in (Balali and Golparvar-Fard 2014). While all these methods have shown reasonable accuracies, their performance has not been benchmarked and compared in the literature. More importantly their application in the context of U.S. traffic signs has never been investigated before.

2.4. Comprehensive Dataset

Arguably, the most pressing challenge with research on detection and classification of US traffic signs is the lack of public image databases to train and test new algorithms. Currently, every publication uses a different dataset for testing the performance of their algorithms which makes benchmarking and comparison of these methods very difficult. To standardize the research problem, several European research projects have released public datasets of traffic signs. Examples include the German Traffic Sign Recognition Benchmark (GTSRB) (Stallkamp et al. 2011; Stallkamp et al. 2012), the Swedish traffic signs dataset (Larsson and Felsberg 2011), and the KUL Belgium traffic signs dataset (Timofte et al. 2014). Nevertheless, to the best of our understanding, there is no comprehensive public databases on US traffic signs which exhibit different visual characteristics compared to their European counterparts.

Because traffic signs are not standardized across different countries, and to save time and effort in data collection and training process, several projects have developed synthetic datasets. For example, (Overett et al. 2011) presents a synthetic dataset which includes several non-US traffic signs. More recently (Mogelmose et al. 2012) trained a sign detection and classification algorithm using synthetic training images, while the testing was conducted on real-world images. However the results from these efforts indicate that the application of synthetic datasets may not be a good solution. While synthetic data can covers a larger variation in the training datasets, they typically do not realistically model real-world variations in traffic sign color, texture, and illumination conditions.

2.5. Data Mining and Visualization

In recent years many data mining and visualization methods are developed that analyze and map spatial data at multiple scales for roadway inventory management purposes (Ashouri Rad and Rahmandad 2013). Examples are predicting travel time (Nakata and Takeuchi 2004), managing traffic signals (Zamani et al. 2010), traffic incident detection (Jin et al. 2006), analyzing traffic accident frequency (Beshah and Hill 2010; Chang and Chen 2005), and integrated systems for traffic information intelligent analysis (Hauser and Scherer 2001; Kianfar and Edara 2013; Wang et al. 2009). (Li and Su 2014) developed a dynamic sign maintenance information system using mobile mapping system (MMS) for data collection. (Mogelmose et al. 2012) discussed the application of traffic sign analysis in intelligent driver assistance systems. (De la Escalera et al. 2003) also detected and classified traffic signs for intelligent vehicles. Using these tools, it is now possible to mine spatial data at multiple layers (i.e. CartoDB) (de la Torre 2013) or spatial and other data together (i.e. GeoTime for analyzing spatio-temporal data) (Kapler and Wright 2005). (Creusen and Hazelhoff 2012) visualized detected traffic signs on a 3D map based on GPS position of the images. (Zhang and Pazner 2004) presented an icon-based visualization technique designed for co-visualizing multiple layers of geospatial information. A common problem in visualization is that these method require adding a large number of markers to a map which creates usability issues and the degraded performance of the map. It can be hard to make sense of a map that is crammed with markers (Svennerberg 2010).

2.6. Retro-Reflectivity Condition Assessment

The most recent mobile retro-reflectivity methods include: **SMARTS** (Sign Management And Retro-reflectivity Tracking System) (Smith and Fletcher 2001), **AMAC** (Advanced Mobile Asset Collection) (Pike and Carlson 2013), **MANDLI** (Retro View) (Harris 2007; Li 2008), and **VISULISE** (Visual Inspection of Signs and Panels) (Evans et al. 2012).

SMARTS was developed by the Naval Research Laboratory for the FHWA. SMARTS include a xenon flash, a laser range finder, one color camera and two monochrome cameras. The unit first sets off a flash and takes a digital image. Then the image is processed to estimate the retro-reflectivity of signs (Rasdorf et al. 2009; Retterath and Laumeyer 2004). This system was an experimental concept and it is not available today. AMAC uses artificial vision and an advanced lighting system to locate, collect, and analyze traffic sign data at night. This data includes: retro-

reflectivity, luminance, position, dimensions, and color. AMAC integrates high accuracy GPS with an onboard inertial navigation system to locate and process sign data (Pike and Carlson 2013). MANDLI continuously fires a high-intensity flash and grayscale cameras simultaneously capture frames. One low intensity camera and one high intensity camera are coupled to cover a wide dynamic range. Flashlights fire infrared light that is invisible to human eyes at a rate of two per second while vehicle travels at highway speed (Harris 2007; Li 2008). VISULISE uses an infrared light. It captures reflected light using a stereoscopic system made up of two high-resolution cameras (Evans et al. 2012). While these methods are practical, yet their application is still costly and in most cases still require operation at nighttime.

Measuring retro-reflectivity from images taken during the day can address the safety concerns. However, several properties needed for retro-reflectivity measurements cannot be directly captured through conventional imaging techniques. Instead, they can be derived by processing pairs of images that are taken in carefully adjusted conditions. For example, image depth can be estimated using two images that are taken with some disparity (Lazaros et al. 2008). Motion can also be estimated using two time-lapse images that are taken from the same location (Balali and Golparvar-Fard 2015; Horn and Schunck 1981).

Reflection can also be estimated by analyzing two images that are captured with different polarizations (Chen and Wolff 1998). By capturing two images through the haze with different polarization filtering – a technique that has been used long in photography- (Schechner et al. 2003) demonstrated a technique that improves visibility. Other improvement technique have also been developed: (Agrawal et al. 2005) presented a technique to remove flash photography artifacts by processing images taken from different flash exposures. This technique improves image quality by removing unwanted specular reflection from the camera flash. (Raskar et al. 2004) developed a technique to capture and convey shape features from real-world scenes. They used a camera with carefully place flashes to detect depth discontinuities and distinguish them from intensity edges due to material discontinuities.

Our technique is similar to that of (Raskar et al. 2004) as we also use a carefully designed artificial light source to capture the physical aspects of the scene. Different from (Raskar et al. 2004) which is designed toward extracting edges and shapes from the image, we estimate a retro-reflectivity map. To do so, first an understanding of luminance and illuminance is needed. Luminance is perceived by the human viewer as the brightness of a light source (Hiscocks and

Eng 2011). Gladly, a pixel intensity value in an image taken with a digital camera is proportional to the luminance in the original scene. This strategy eliminates the need for expensive luminance meters, and has the following advantages (Wüller and Gabele 2007):

- Each of the millions of pixels in the CMOS sensor of a camera becomes a luminance sensor, and thus a digital camera can capture the luminance of an entire scene. This speeds up the measuring process and allows multiple measurements at the same instant.
- The surroundings of the luminance measurement are recorded which puts the measurement in its context.
- For luminance measurement, the field of view (FOV) of a visual sensor must be smaller than the size of the light source. The FOV of a digital camera pixel is on the order of 150 times smaller than the FOV of a luminance meter which is about 1°. Thus, the CMOS sensor of a camera is powerful enough to measure small area light sources such as individual light emitting diodes. These light sources are difficult or impossible to measure with a luminance meter (Hiscocks and Eng 2011).

By photographing a source of known luminance and calibrating the camera, we obtain the conversion factor that links luminance (in candela per square meter($^{cd}/_{m^2}$)) to the intensity value of a pixel in an image. This creates the basis for our method which is presented in Chapter 3.

Over all, there is a need for improvement on the state-of-the-art vision-based techniques for 3D reconstruction, detection and localization of assets. Particularly, a new environment needs to be created where the geometrical (3D) and appearance (2D image) information of the assets is integrated, providing a platform for development of joint recognition and localization of assets. In the following, our new method for creating such an integrated environment, plus joint 3D reconstruction and segmentation of assets from 3D point clouds is presented in detail.

CHAPTER 3. METHOD

The main focus of this research is to test whether the hypothesized computer vision based framework shown in Figure 3.1 can detect, classify, and spatially locate low-cost high volume roadway assets specifically traffic signs from an array of cameras mounted on a road inspection vehicle. Automated recording of the types, locations and up-to-date status of the civil infrastructure assets enables state and local transportation agencies to plan, design, construct, operate, and manage their transportation systems more effectively; eliminates the need for labor-intensive, time-consuming, costly and unsafe manual or semi-automated practices, and finally supports development of automated condition assessment and context-aware operation and maintenance applications.

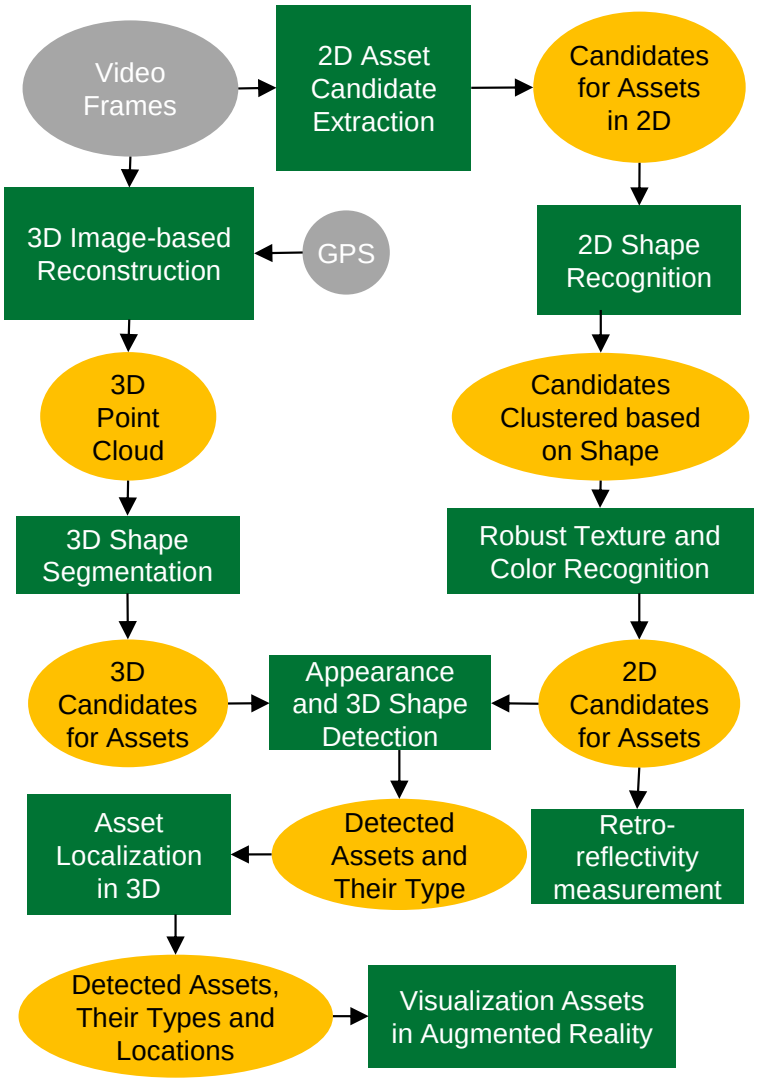

Figure 3.1 Computer Vision-based Research Framework

In the proposed hypothesized framework, a computationally effective image-based 3D reconstruction algorithm takes the video frames and reconstructs a dense 3D point cloud model of all visible objects. Using a new 3D shape segmentation algorithms, 3D points are hierarchically clustered to form a potential set of assets. For each candidate, a feature vector based on joint representation of shape and context is formed and is fed into another SVM classifier. This algorithm classifies 3D assets such as guardrails and light poles which are not detectable from a 2D video frame. The selection of 2D and 3D candidates for assets is further refined by using a novel joint appearance and 3D shape recognition classifier.

In the meantime, using the video streams collected from the cameras mounted on the vehicle, a set of thresholded frames is initially identified. Each thresholded frame contains a set of candidates for assets along the right side of the roadway (e.g., mile markers, traffic signs). Using a new Support Vector Machine (SVM) classifier and based on the color channels at pixel level, a set of bounding boxes are initially extracted wherein each bounding box potentially includes an asset. While this algorithm returns very few FNs (e.g., non-detected assets), it is purposefully designed to return a high number of FPs (e.g., potential candidates for assets). This stage passes all assets that are partially occluded (e.g., behind a tree or a parked vehicle), damaged or their signage is faded (e.g., faded stop sign). Next, using a new shape recognition algorithm based on Haar-like features, the 2D candidates are further refined and categorized based on their shape appearances (e.g., rectangle, diamond). These candidates are placed into a new texture and color recognition algorithm, wherein by using a multiple binary SVM classifier, they are further classified into particular predefined types of assets. This step is mainly detecting traffic signs and mile markers which are recognizable in 2D images. Finally using connectivity semantics embedded between the video frames and 3D points in the reconstructed point cloud, assets are localized in 3D. The detected assets and their types are visualized in an augmented reality environment which enables remote walk-throughs for inventory management. The semantically-rich augmented reality environment also serves as a map for operation and maintenance context aware applications using smartphones and tablet PCs.

The motivations behind the proposed computer vision based framework lie in the deficiencies of the current practice of infrastructure asset management, and the transformative potential of using mounted cameras on an inspection vehicle as sensors and reporters of the location and up-to-date status of the assets and their conditions.

3.1. Segmentation and Recognition of Roadway Assets Using Image-based 3D Point Clouds and Semantic Texton Forests

Given a collection of video frames collected from a car-mounted camera, the goal is to:

a) Reconstruct a 3D point cloud model of the roadway assets;

b) Segment the 2D images at the pixel level into several categories of assets;

c) Color-code and label the reconstructed 3D point cloud model based on the detected asset categories from the 2D segmented images.

The proposed approach is illustrated in Figure 3.2.

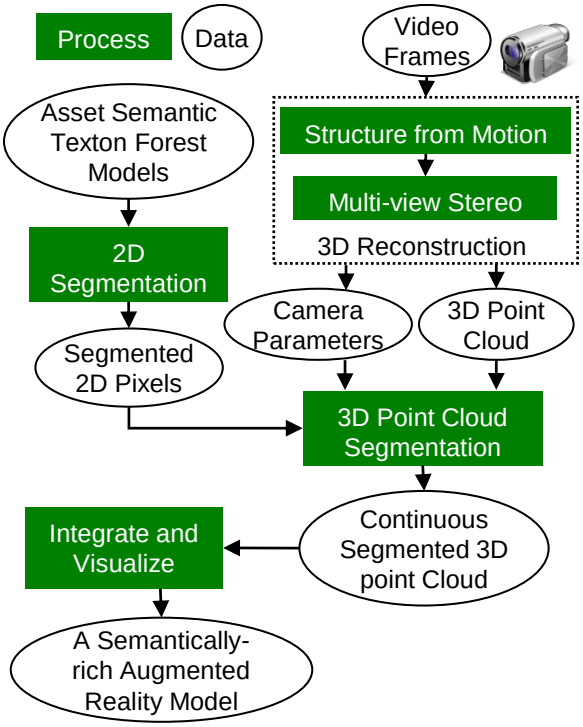

Figure 3.2 Flowchart of the Proposed Segmentation and Recognition Approach

It is assumed that each asset is at least visible from a minimum of three video frames. These frames contain typical dynamic roadway foregrounds and backgrounds and can include sky. In the training stage of our proposed 2D segmentation method, it is assumed that for each video frame, a ground truth image is carefully generated in which the parts of the image that correspond to the asset categories are labeled and color-coded accordingly. For this purpose, a comprehensive image dataset for 12 different types of asset categories is created. In our dataset, each image can contain more than one type of asset. For these images, the ground truth is labeled and color-coded for all

observed types of assets, and then the image is used in all appropriate corresponding training categories.

Using an improved image-based 3D reconstruction pipeline consisting of Structure from Motion (SfM) and Multi View Stereo (MVS) algorithms (Golparvar-Fard et al. 2012), a point cloud model of the roadway and all assets along is reconstructed and the images are geo-registered in a common 3D environment. For each image, our proposed algorithm for 3D point cloud segmentation uses both textons and priors as features to give coherent semantic segmentation and labels each pixel with an asset category accordingly. In the 2D segmentation process, inspired by the bag of semantic textons (Shotton et al. 2008), the image is categorized into the asset categories; i.e., the image is simultaneously segmented into coherent regions and each region is categorized accordingly. Based on consistent geometrical correspondence among labeled pixels from the underlying 3D point cloud model and a multi-class scoring mechanism, the corresponding 3D points are labeled for the highest classification score. The resulting segmented and geo-registered imagery along with the point cloud are visualized in a common 3D environment. In the following, each step is discussed in detail:

3.1.1. Image-based 3D reconstruction

3D image-based reconstruction pipeline which is an algorithmic improvement to (Uslu et al. 2011) consists of two steps that are performed sequentially:

- Structure from Motion (SfM) which helps generating a sparse 3D point cloud model and calibrate all uncalibrated video frames;
- Multi-view Stereo (MVS) which takes the calibrated cameras, and generate a dense 3D point cloud model.

Compared to (Uslu et al. 2011), several components of our new pipeline are implemented on Graphic Processing Unit (GPU) or use multi-core CPU. This has significantly reduced the computational time which is necessary when dealing with long sequences of video streams. Figure 3.3 shows an overview of our image-based 3D reconstruction algorithm.

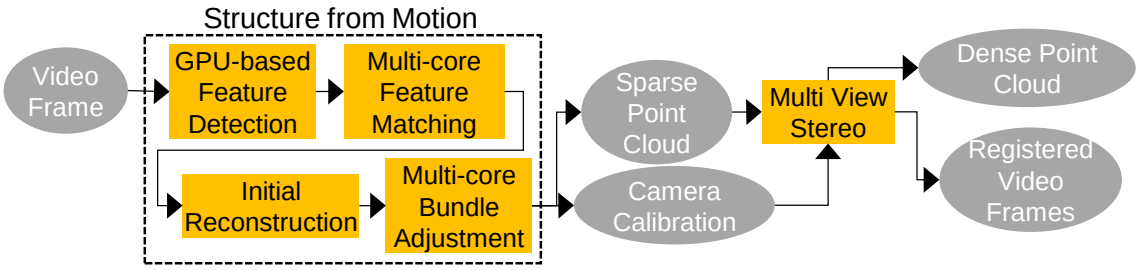

Figure 3.3 Flowchart of Image-based 3D Reconstruction

In the SfM algorithm, first visual features are independently extracted for each video frame. In our new approach, we use Scale Invariant Feature Transforms (SIFT) that is implemented on GPU (Wu 2007). Next, using a new multi-core implementation, the SIFT features are matched in pairs over the span of Ω consecutive video frames. An initial solution for the 3D locations of these features points is calculated using Nister's 5-point algorithm (Nistér 2004). The camera's parameters are calculated using the Direct Linear Transform (DLT) technique (Hartley and Zisserman 2003) inside a RANSAC procedure (Fischler and Bolles 1981). The DLT also gives an estimate of the intrinsic parameter matrix. For those video frames for which their matching feature points give a well-conditioned estimate of their locations (Golparvar-Fard et al. 2012) the video frames are incrementally added, until no remaining camera observes any reconstructed 3D point. The objective function for the distance between SIFT features and their re-projected 3D points at every iteration is minimized through an optimization process using the multi-core sparse bundle adjustment library of (Wu et al. 2011). This process results in a sparse point cloud model, plus intrinsic and extrinsic camera parameters for each video frame which are fed into the MVS algorithm (Furukawa et al. 2009) to improve density of the sparsely reconstructed model.

The MVS algorithm (Furukawa et al. 2009) consists of a match, expand, and filter procedure. At first, during the matching phase for all calibrated cameras, a set of new features using Harris and difference-of-Gaussians operators are found and matched across multiple video streams, yielding a sparse set of patches associated with salient video frame regions. Given these initial matches, the following two steps are repeated:

- Expansion: which spreads the initial matches to nearby pixels and obtain a dense set of patches;
- Filtering: which eliminate incorrect matches using visibility constraints.

Similar to (Furukawa et al. 2009), our implementation of the algorithm replaces their greedy expansion procedure by iteration between expansion and filtering steps, which processes complicated surfaces and rejects outliers more effectively. The output from consecutive SfM and MVS steps results in a dense 3D point cloud and geo-registers all video frames into the same coordinate system.

3.1.2. 2D segmentation

Our 2D segmentation and asset recognition method is primarily inspired by Semantic Texton Forests (STFs) (Shotton et al. 2008). STFs are powerful low-level features which are employed for the semantic segmentation of 2D video streams based on different asset categories. Since this approach directly acts on video stream pixels, it does not need the expensive computation of commonly used filter-bank responses or local descriptors. Without performing time-consuming K-means clustering and nearest neighbor assignments, STFs enable powerful texton codebooks to be built. Due to their superior quantitative performance and execution speed over other algorithms, STFs are chosen for 2D semantic segmentation of roadway assets from long sequences of video streams.

The STF algorithm in our work contains randomized decision forests that use only simple pixel comparisons on local image regions, performing a hierarchical clustering into semantic textons and a local classification of the asset region category. Here, the randomized decision forests are used as a machine learning technique to categorize individual pixels of the video frames into the most appropriate asset categories. A randomized decision forest combines the output of many different decision trees, each of which has a different structure and split tests. The term randomized refers to the procedure of the training algorithm as:

- Each tree is trained on a random subset of the roadway asset data,
- When the trees are being built, several candidate split tests are chosen at random from a large pool of potential features.

The test that optimally splits the data is taken under an optimization criterion chosen at the training stage. As validated in (Johnson and Shotton 2010), these two forms of randomization ensure that no two trees in the forest can over-fit to the whole training set.

Our goal of using the decision forests is to determine the asset category *c* of a pixel *p*, given the context around that pixel. Here, the context refers to the surrounding of the roadway assets

which appear in a 2D image. In our work, we assume we have a supervised labeled training dataset; i.e., our training dataset is manually labeled for ground truth. Each forest contains trees with nodes *n*, and leaf nodes *l*. Associated with each node is a learned asset category distribution *P(c|n)*. An example semantic texton tree is illustrated in Figure 3.6, in which a tree has been trained on pavement marking and asphalt images and can effectively segment an image according to these two semantic asset categories. The whole forest achieves an accurate asset classification once new pixels are being chosen by averaging the class distributions over the leaf nodes $l_p = (l_1, l_2, \dots, l_T)$ reached by the pixel *p* for all *T* trees:

$$P(c \mid L(p)) = \sum_T P(c \mid l_t) P(t) \tag{4.1}$$

An example of the overall structure of the semantic texton forest is shown in Figure 4.4. Each tree in the forest is built separately on a subset of the training images.

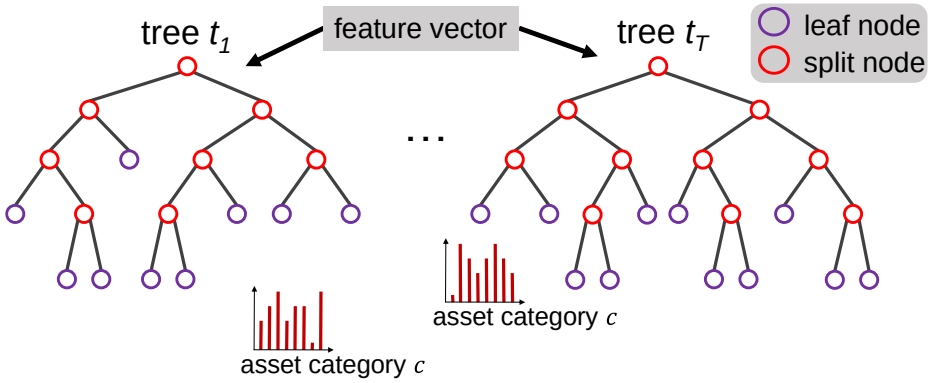

Figure 3.4 Decision Forests (Inspired by (Shotton et al. 2008))

The training data consists of a set *P* of pixels sampled from training images and ignoring pixels marked as background. To ensure good estimates of the tree class distributions, the entire set of pixels is used later to fill the tree after construction.

Each decision tree is constructed by partitioning *P* into two subsets P_{left} and P_{right} based upon a split test. P_{left} is used to create the left sub-tree and P_{right} is used for the right one. This process is repeated until a thresholding condition is met. The split test used to partition *P* is chosen in the same manner as (Johnson and Shotton 2010; Lepetit et al. 2005); by examining STFs and the possible tests and selecting a combination that maximizes the expected information gain about the node categories. The information gain is calculated as:

$$\Delta E = -\frac{|P_{left}|}{|p|} E(P_{left}) - \frac{|P_{right}|}{|p|} E(P_{right}) \quad (4.2)$$

where *E(I)* is the Shannon entropy of the classes in the set of example pixels *P* (Shotton et al. 2008).

a. Training a randomized decision forest

Similar to (Johnson and Shotton 2010), training a randomized-decision forest for the asset categorization involves choosing several parameters which are as follows:

- ***Number of Trees in the semantic texton forest.*** Which is adjusted based on both accuracy and computational time of the 2D asset segmentation process.
- ***Type of Split Tests.*** This can have a significant role in the performance of training as various split tests can work in a complimentary form.
- ***Maximum Tree Depth.*** There are different kinds of pixel tests which can be implemented. Deeper decision trees result in better asset segmentation; in the meantime, it will dispose the trees to the over-fitting problem.
- ***Window Size.*** The dimensions of the window around each pixel can impact both local features and contextual information. Larger windows produce more features but they are less likely to develop to other pixels.
- ***Information Channels.*** The quality and number of information channels and how they are used for asset segmentation can impact both application of STF method and also what the type of tests are selected. For instance for color feature in asset images, the methods which use the color at pixel-level are formed.

The selection process for these parameters is a function of the characteristics of roadway asset dataset and their features. Small asset dataset will need superficial trees and typical larger dataset will need deeper trees. For finding the best combination of parameters for asset category segmentation, several experiments were conducted in this study. The pixel tests which are used in the experiments have been listed in Table 3.1. $P[w_0]$ and $Q[w_1]$ are the values of pixels within a patch of size $q \times q$ centered on the training pixel (See Figure 3.5). It is not necessary for the channels w_0 and w_1 to be exactly the same. As long as a test such as the difference and absolute difference of pixels is equal to a global intensity shift, others are just possible discriminative combinations of

pixels. In addition to these pixel tests, the Haar-like features of (Viola and Jones 2001) and the rectangle sum features of (Shotton et al. 2009) have been used.

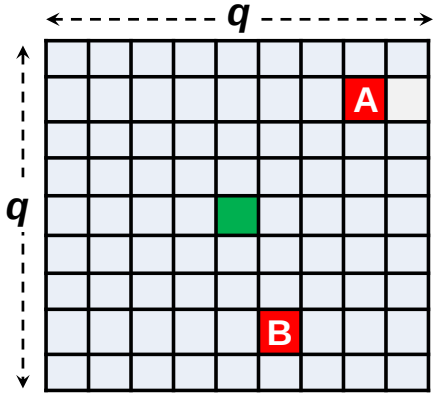

Figure 3.5 Pixel Comparison Split Test

Table 3.1 Split Tests Based on Image Information

Label	Test Domain
1	$P[W_0]$
2	$Log(P[W_0])$
3	$P[W_0]+Q[W_1]$
4	$P[W_0]-Q[W_1]$
5	$\|P[W_0]-Q[W_1]\|$
6	$P[W_0]log(Q[W_1])$
7	$P[W_0]\times Q[W_1]$
8	$P[W_0]/Q[W_1]$

Our supervised labeled image data is then used to train a semantic texton forest, consisting pairs *(p,c)* of pixels *p* and asset category *c* labels. Consequently, each pixel is given a training label as shown in Figure 3.6. During training, the distribution *P(c|n)* is computed as a normalized histogram of the training tuples which reached a particular node *n*:

$$P(c \mid n) = \frac{H_n[c]}{\sum_c H_n[c]} \tag{4.3}$$

where $H_n[c]$ is the number of pixels of asset class *c* that passed through a node *n* during training. Filling, the process of computing this histogram at each node is performed using all of the pixels in the training data, by passing each pixel down a tree and incrementing the relevant histogram bin $H_n[c]$.

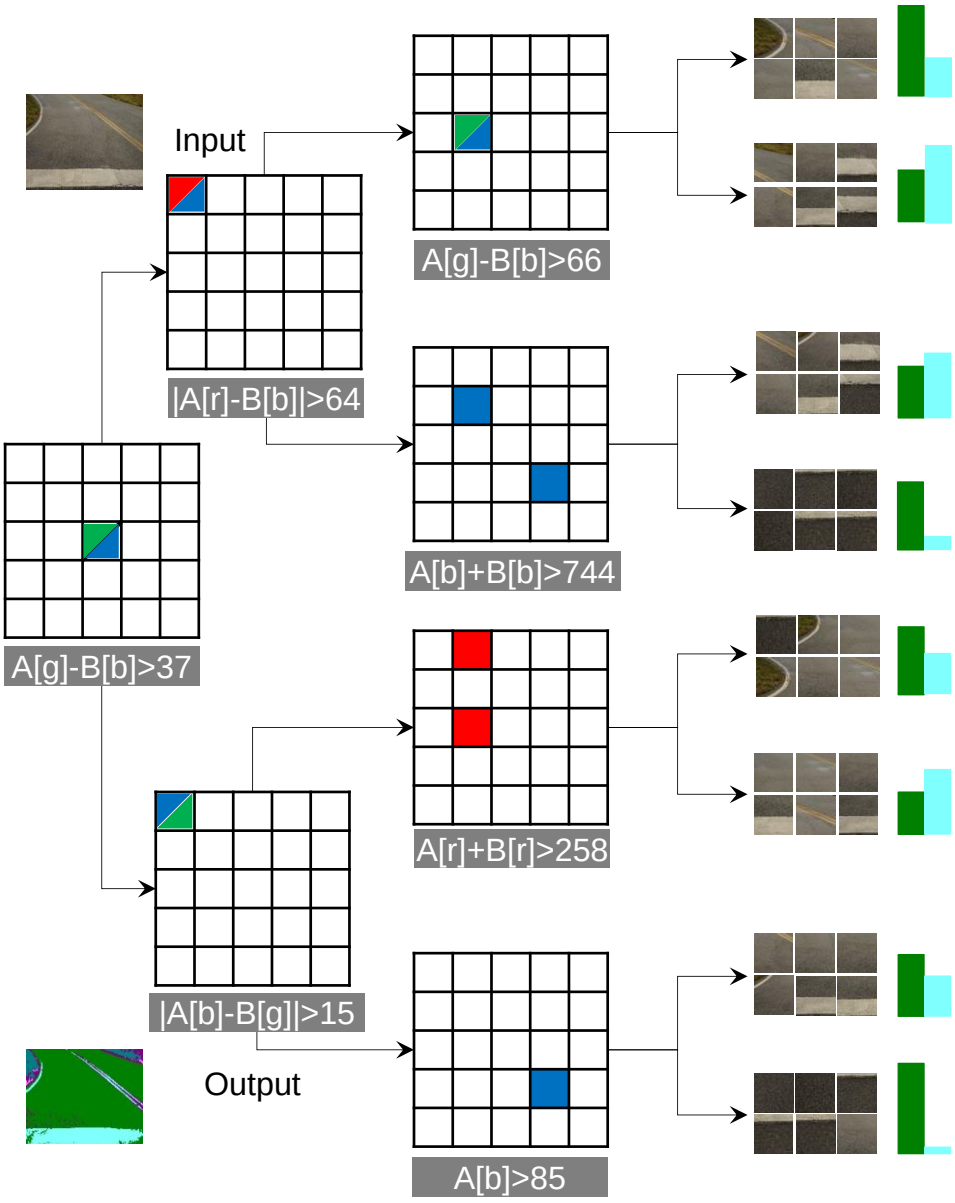

Figure 3.6 An Example of a Semantic Texton Tree for Assigning an Asset Label to a Pixel

b. Asset categorization over an image region using bag of semantic textons

In order to categorize the image for different types of assets, we use the bag of semantic textons which combines a histogram of semantic textons over an image region with a region prior category distribution. First, a non-normalized histogram $H_r(n)$ that concatenates the occurrences of tree nodes n across the different trees is formed. A conditional distribution over the region given by the average class distribution is also computed:

$$P(c \mid n) = \sum_{p \in r} P(c \mid L_p) P(p) \tag{4.4}$$

Experiments are performed with tree histograms where both leaf nodes *l* and split nodes *n* are included in the histogram, such that:

$$H_r(n) = \sum_{n' \in child(n)} H_r(n') \tag{4.5}$$

This histogram therefore uses the hierarchy of clusters implicit in each tree. Each *P(c|L(p))* is already averaged across trees, and hence there is a single region prior *P(c|r)* for the whole forest. Figure 3.7 shows the bags of semantic forest.

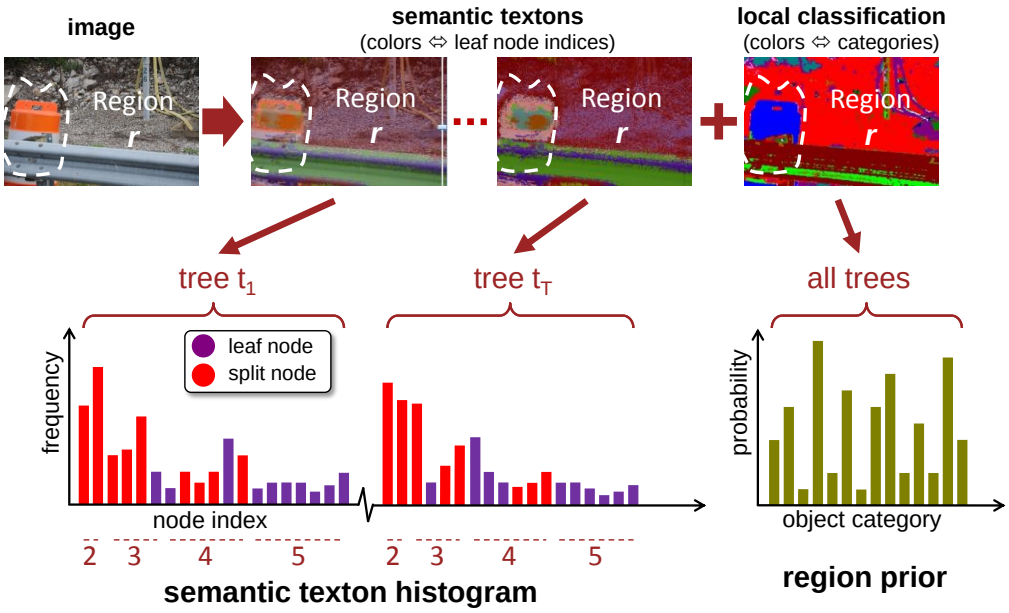

Figure 3.7 Semantic Texton and Region Prior Histograms

3.1.3. Segmentation and recognition of assets from 3D point clouds

Single-view recognition is just a preprocessing stage, and the final decision results from global optimization over multiple views. Given a set of consecutive video frames that observe a region in 3D, single-view recognitions, camera positions and calibrations, our algorithm votes for a possible set of 3D hypotheses for different roadway asset categories. The category which returns the maximum vote per reconstructed point 3D will be the final outcome of the process. In practice, for each reconstructed 3D point *(P_i)* that is observed from multiple cameras, a normalized

histogram is formed which captures the relative frequency of the times different 2D segmentation category *(Cj)* with $j \in \{1,2,...,k\}$ are observed from these cameras. Once these histograms are formed, the categories with the maximum voting from corresponding video frames (or the asset category bin in the histogram with the highest frequency rate) will be assigned to the 3D points (See Figure 3.8). In other words, if a set of semantic texton labels satisfies consist geometrical and visual constraints, then all of these labels are explainable by one 3D asset category. Figure 3.9 shows the corresponding video frames that generate the asset label for the 3D points in the reconstructed cloud.

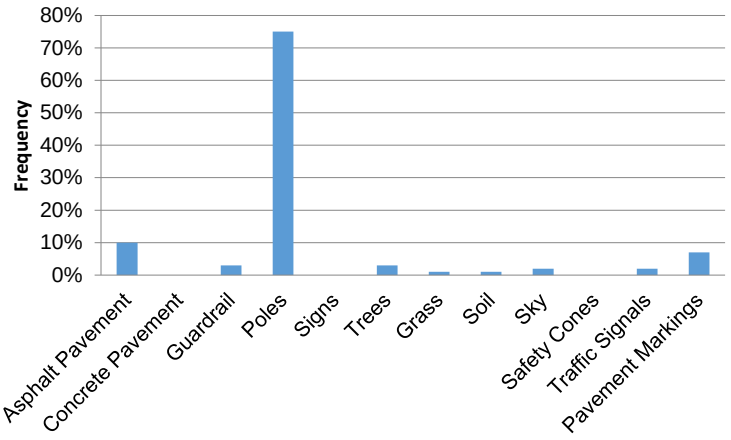

Figure 3.8 Histogram for Labeling a 3D Point in the Reconstructed Cloud: The Category Returning the Maximum Frequency of Appearance Across All 2D Imagery That Observes the 3D Point Will Be Chosen

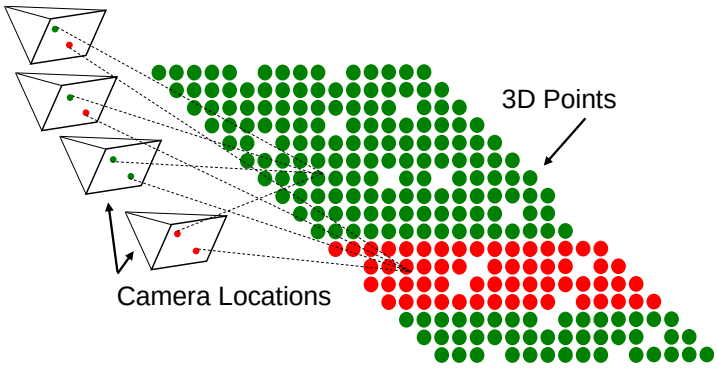

Figure 3.9 Semantic Labeling of the 3D Points in the Cloud Voted Based on the Labels of the Corresponding Image Pixels

3.1.4. Visualization module

In our visualization platform, in addition to rendering the 3D point cloud model, the digital images (or the locations of the camera) are also rendered in form of pyramid-shape camera frusta.

Once a camera frustum is visited by the user in the reconstructed 3D scene, the frustum is automatically texture-mapped with the full-resolution images. Here, the user can either select to view the original digital image or the 2D segmentation results. The same holds true for the 3D point cloud, as both original colored point cloud, and also the results of the 3D labeling can be shown. The user location can also be preserved, while the viewpoint is changing to jointly visualize and study the 3D point cloud models and the geo-registered imagery. Also the user has the ability to filter one type of asset of interest and only focus on the related parts in 3D and their geo-registered imagery. This can minimize the amount of time a user has to spend to navigate through all existing photos and find areas of interest for condition assessment purposes.

While this section of my research presented the initial steps towards processing site video streams for the purpose of roadway asset categorization, several critical challenges remain. Some of the open research problems include:

- ***Recognition of roadway assets in long video sequences.*** The presented algorithm proposes a pipeline for efficient segmentation and 3D reconstruction of roadway assets. Practitioners can use the results of these segmentations to quickly and easily navigate through imagery and identify particular categories of assets. Yet, this work primarily segments a point cloud into different categories of assets, and does not fully recognize and classify different types of assets. Furthermore, it does not distinguish the intra-class variability in assets which is a key component in asset data collection and condition assessment; e.g., stop sign vs. speed limit sign.
- ***3D localization of roadway assets.*** So far, the user can localize assets in a supervised fashion; i.e., practitioners should select certain areas from 3D or their corresponding 2D regions to extract the location of the assets in 3D. More work needs to be done on integrating asset detection algorithms such as (Timofte et al. 2014) and (Hu and Tsai 2011) with the presented work for automated localization purposes.
- ***Need for semi-supervised segmentation techniques.*** One of the major challenges with the proposed algorithm is its reliance on supervised ground-truth pixel labels. Generating such data is time-consuming and labor-intensive. Moreover, due to the large dimensionality of the bag of semantic textons, increasing the size of the data would increase the computation time. Hence, there is a need for unsupervised or semi-supervised techniques that can provide ground truth data in a more efficient manner.

3.2. Segmentation and Recognition of Roadway Assets from Car-Mounted Camera Video Streams using a Scalable Non-Parametric Image Parsing Method

Our method, as shown in Figure 3.10, obtains video frames and semantically and geometrically labels different parts of the video frames into roadway asset categories such as guardrail, light poles, traffic signs, pavement, and etc. The method builds on superparsing algorithms which are known as simple and effective nonparametric methods for labeling image regions into certain object categories (Balali and Golparvar-Fard 2014; Liu et al. 2011; Tighe and Lazebnik 2013).

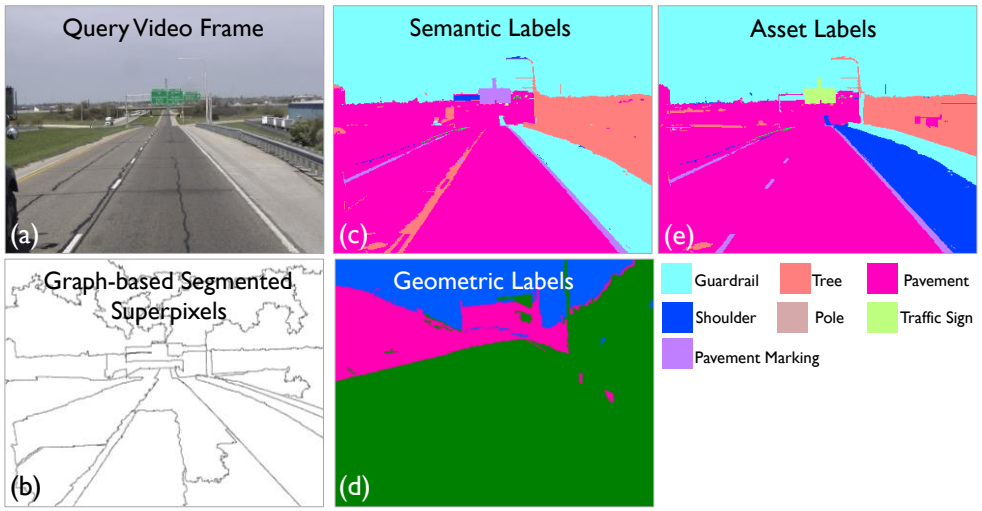

Figure 3.10 System Overview for Segmentation Process

Here, we have a relatively large number of asset categories, while we have to be dealing with huge video databases that are already being collected. Manual and very time-consuming process for preparing the training data (ground truth) for such large number of assets from very large video datasets which contain all types of scenes, with different levels of clutter, occlusion, and lighting/environmental condition is an important issue Thus, instead of a time-consuming (completely manual) fully supervised training process, we leverage a lazy scheme for training the model. In artificial intelligence, ”lazy” is used for learning method in which generalization beyond the training data is delayed until a query is made to the system. It simply stores training data (or only minor processing) and waits until it is given a test image. It means that almost no training takes place offline and as a result takes less time in training but more time in predicting the labels of asset categories. Given a test image to be segmented and labeled, the method dynamically selects a group of training samples that appear to be the most relevant and proceeds to transfer the

labels from selected training images to the test image. The training in essence is a semi-supervised machine learning algorithm which predicts and assigns asset labels from little labeled and a lot of unlabeled data. One natural way, would be to assign asset labels per pixel; nevertheless as stated in the previous section, our prior work and others (Golparvar-Fard et al. 2012; Shotton et al. 2008) show assigning labels at the pixel-level tends to be computationally costly and inefficient. In order to increase efficiency, the labels are assigned to superpixels. Superpixels are 2D image regions that are produced from pixels through a fast graph-based segmentation algorithm (Felzenszwalb and Huttenlocher 2004; Malisiewicz and Efros 2008). Not only does this strategy reduce the complexity of the asset categorization, but also gives better spatial support for assembling visual features that could belong to a single roadway asset. This is because there will be no need for fixed size square patches around each pixel for labeling purposes.

Once the superpixels are obtained from each video frame, their appearance is described using a bank of superpixel filters. Having extracted the superpixels along with their features, a likelihood ratio score is obtained for each superpixel and independent semantic labels (e.g. guardrails) and geometric labels (e.g. horizontal) are assigned. To do so, the superpixel scores are matched with a relatively small set of training images that serve as the source of annotations for the superpixels (denoted as "Retrieval Set of Training Images" in Figure 3.11). Choosing a smaller dataset is done to minimize the computational time and provide scene-level context for subsequent superpixel matching steps; for example by using roadway images in the database vs. secondary road images. To quickly choose the subset of training images, we only extract four global features for the superpixels (different from those used in the bank of superpixel filters above). Based on their Euclidean distance, these global features are ranked against their counterparts from the training images. We then only take the top *k* images – the Retrieval Set– as the representative of the training sample images. The matching of the superpixel scores with the training images is done using a standard Markov Random Field (MRF) optimization. Here the semantic and geometric labels are assigned simultaneously. Be enforcing coherence between the geometric and semantic labels, we then improve the accuracy of the semantic labeling and produce the "asset labels". Figure 3.11 shows the data and process steps in our method. The first row in Figure 3.12 illustrates the query image and the retrieval set. The second row shows an example of the ground truth geometric and semantic labels.

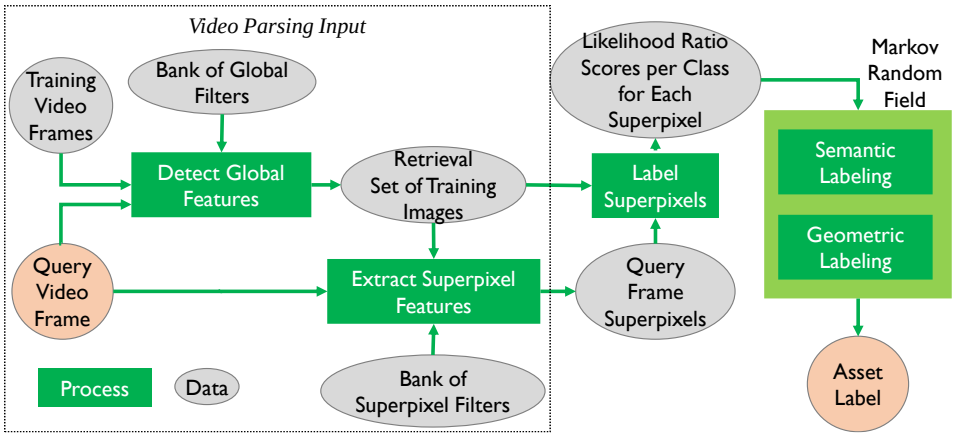

Figure 3.11 Overview of the Video Frame Segmentation Process

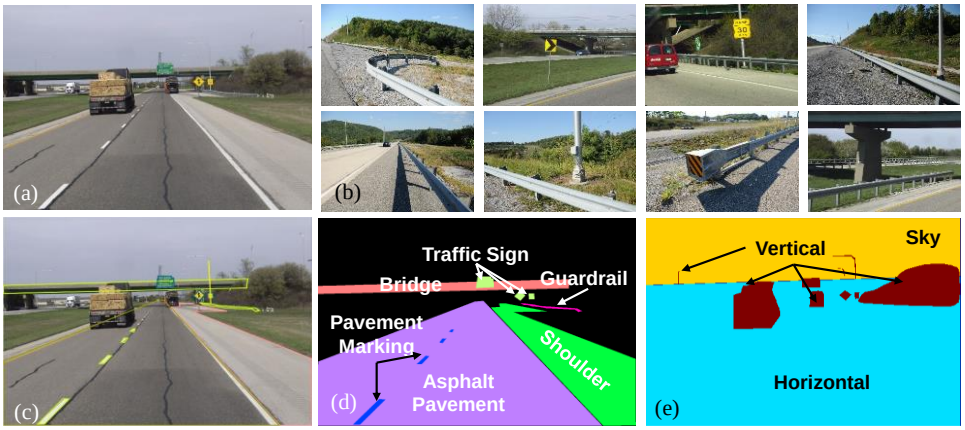

Figure 3.12 Training Process: (a) Query Image; (b) Retrieval Set of Similar Images; (c) Labeled Images Using LabelMe Toolbox; and Ground Truth for (d) Semantic Labels; (e) Geometric Labels

3.2.1. Retrieval set of training images

The first step is to find a small retrieval set of training images that will serve as the source of candidate superpixel-level matches. This is done for computational efficiency, and also providing scene-level context for the subsequent superpixel matching step. A good retrieval set will contain images that have similar road scenes, high-quantity low-cost roadway assets, and spatial layouts to the query image (e.g. secondary road vs highway, urban vs. country roads). Three types of global features are used to capture this kind of similarity: spatial pyramid, gist, and color histograms (Lazebnik et al. 2006; Oliva and Torralba 2006; Tighe and Lazebnik 2013). Spatial pyramid is a collection of order-less feature histograms computed over cells defined by a multi-level recursive image decomposition. The gist is an abstract representation of the scene that spontaneously activates memory representation of scene categories (e.g. a city, a mountain). Color

information is also used to capture the similarity between roadway assets. Table 3.2 shows the specification of these global features in detail.

Table 3.2 Global Features for Retrieval Set Computation

Feature	Type	Description	Dimension
Spatial Pyramid	3 levels	SIFT dictionary size 200	4200
Gist	3 channel RGB	3 scales with 8, 8, and 4 orientation	960
Color Histogram	3 channel RGB	8 bins per channel	24

The retrieval set is the source of possible labels and region-level matches. The appropriate size of retrieval set depends on a size of dataset and on the distribution of the asset categories contained in them. While the total number of asset categories is high, a single image only contains a small subset of all possible assets. For each feature type, all the training images are ranked in an increasing order of Euclidean distance from the query image. Then the minimum rank of each feature is used to get a single ranking for each image and use the top ranking *k* images as the retrieval set. As a result, after conducting preliminary experiments, we used the size of 200 for both of these datasets. This also minimizes the number of memory-to-disk input/output times on the descriptors for each test image which can slow down the process for larger sizes of the retrieval set.

3.2.2. Superpixel features

The superpixels are obtained using the fast graph-based segmentation algorithm of (Felzenszwalb and Huttenlocher 2004) and described using twenty different features similar to (Malisiewicz and Efros 2008) that encode shape, location, texture, color, and appearance information (see Table 3.3). All of these features are computed for each superpixel in the training images and stored together with their asset category. The asset category is associated with a training superpixel if more than 50 percent of the superpixels overlaps with ground truth segment mask with that asset label. Histogram of textons and dense SIFT (Scale Invariant Feature Transform) descriptors are computed over the superpixel region. Textons refer to fundamental micro-structures in natural images (and videos) and are considered as the atoms of pre-attentive human visual perception of a good mathematical model. Textons are a discrete set of representative local features for the objects. The basic idea is to encode efficiently how the textons transform

when illumination, camera pose, and other parameters. Thus, it is a very useful concept in pattern recognition and has been utilized to develop efficient models in the context of texture recognition and object detection. For SIFT features which are more powerful than textons, left, right, top, and bottom boundary histograms are used.

Table 3.3 Features Used for Segmenting the Superpixels

Feature	Type	Description	Dimension
Superpixel Shape	Shape	Masking the shape of the superpixel over its bounding box	8×8
Superpixel Aspect Ratio	Shape	Bounding box width/height related to image width/height	2
Superpixel Area	Shape	Superpixel area relative to the area of the image	1
Superpixel Shape	Location	Mask of superpixel shape over the image	8×8
Top Height	Location	Top Height of bounding box relative to image height	1
Texton Histogram	Texture	Texton histogram, dilated by 10 pixel texton histogram	100×2
SIFT Histogram	Texture	Quantized SIFT histogram, dilated by 10 pixel quantized SIFT histogram	100×2
Boundary SIFT Histogram	Texture	Left/right/top/bottom boundary quantized SIFT histogram	100×4
Color Mean + Standard Deviation	Color	RGB color mean and standard deviation	3×2
Color Histogram		Color histogram (RGB, 11 bins per channel), dilated by 10 pixel color histogram	33×2
Color Thumbnail	Appearance	Color Thumbnail	3×8×8
Grayscale Gist	Appearance	Grayscale gist over superpixel bounding box	320

3.2.3. Computing ratio score

Having segmented the test image and extracted features of all superpixels, a log likelihood ratio score for each superpixel *(Xi)* and each category of roadway asset *(a)* that is included in the retrieval set is calculated. As defined in Equation (4.6), the calculation is based on a Naïve Bayes assumption that given the asset category *a*, features (f_i^k) are independent from one another.

$$L(X_i, a) = \log \frac{P(X_i|a)}{P(X_i|\bar{a})} = \log \prod_k \frac{P(f_i^k|a)}{P(f_i^k|\bar{a})} = \sum_k \log \frac{P(f_i^k|a)}{P(f_i^k|\bar{a})} \quad (4.6)$$

where $(\bar{a})$ is the set of all types of assets excluding (a). Each likelihood ratio $\frac{P(f_i^k|a)}{P(f_i^k|\bar{a})}$ is computed with the help of nonparametric density estimates of features from the required asset categories in the neighborhood (f_i^k).

3.2.4. Markov Random Field (MRF)

Next, contextual constraints are enforced on the image labeling, for instance, a "pavement" label assigned to a superpixel completely surrounded by "sky" is not reasonable. This global image labeling problem is formulated as minimization of a standard MRF energy function defined over the field of superpixel labels $a = \{a_i\}$:

$$J(a) = \sum_{X_i \in \Phi} E_{data}(X_i, a_i) + \lambda \sum_{(X_i, X_j \in \psi)} E_{smooth}(a_i, a_j) \tag{4.7}$$

where Φ is the set of superpixels, ψ is the set of pairs of adjacent superpixels and λ is the smoothing constant. The data term is defined as in Equation (4.8) where $L(X_i, a_i)$ is the likelihood ratio score from Equation (4.6), $\sigma(t) = \frac{\exp(\gamma t)}{1+\exp(\gamma t)}$ and ω_i is the superpixel weight (the size of X_i in pixels divided by the mean superpixel size):

$$E_{data}(X_i, a_i) = -\omega_i \sigma(L(X_i, a_i)) \tag{4.8}$$

3.2.5. Simultaneous classification of semantic and geometric classes

We assume that each semantic class is associated with a unique geometric class and specify this mapping manually in the training process. For example, a light pole is "vertical", a guardrail is "horizontal", and so on. We leverage this constraint to explore a higher-level form of context information for asset labeling purposes, and manifest this in form of simultaneously labeling regions into two types of classes: semantic and geometric. Like (Gould et al. 2009; Tighe and Lazebnik 2013) three geometric labels (sky, horizontal, and vertical) are used although the sets of

semantic labels in our datasets are much larger. A cost function as show in Equation (4.9) is used to jointly infer semantic *(a)* and geometric *(g)* labels, where in (ζ) is the term that enforces consistency between the geometric and semantic labels:

$$H(a,g) = J(a) + J(g) + \mu \sum_{X_i \in \psi} \zeta(a_i, g_i) \tag{4.9}$$

$$\zeta = \begin{cases} 0 & if\ the\,(a_i)\,is\ of\ the\ type\,(g_i) \\ 1 & otherwise \end{cases} \tag{4.10}$$

3.2.6. Video parsing

Motion cues available in videos can improve asset segmentation, as they allow the same asset to be visible from multiple video frames, possibly from various view points and scales. Leveraging such cues can help with a better understanding of the shape and appearance of an asset. To do so, a query video is pre-processed using a hierarchical video segmentation method (Grundmann et al. 2010) that gives 3D region which have roughly uniform color and optical flow. Once the 3D region associated with a video frame is obtained, a data term $E_{data}(\upsilon_i, a)$ for each supervoxel (υ_i) - the volumetric space associated with the frame – and asset category *(a)* is calculated. Then local likelihood scores for possible asset categories over each region is calculated and eventually a single graph MRF for each video frame is constructed where nodes represent supervoxels and edges connect pairs of supervoxels that are spatially adjacent in at least one frame. Figure 3.13 shows how this can improve the proposed method for single video frames in Figure 3.11.

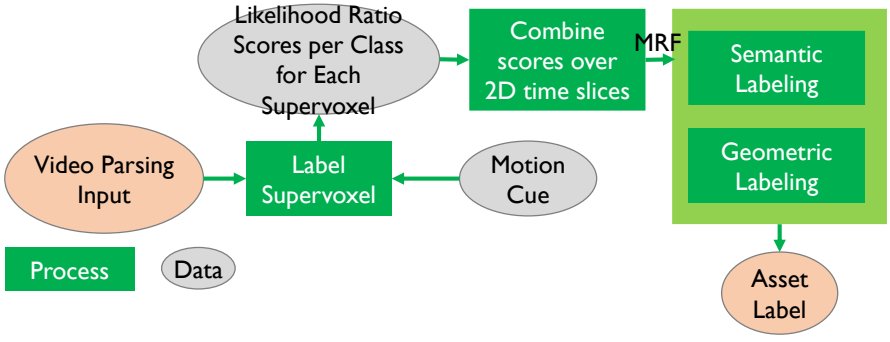

Figure 3.13 Overview of Video Parsing Process (*Video Parsing Input* Comes from Figure 3.11)

3.3. Evaluation of Multi-Class Traffic Sign Detection and Classification Methods for U.S. Roadway Asset Inventory Management

To address the current gaps in the literature, we present and compare the performance of three different methods for detection, 2D localization, and classification of multiple categories of US traffic signs (warning, regulatory, yield, and stop sign) by leveraging both shape and color features. Our dataset is formed from real-world video streams that are collected from the cameras mounted on the DOT inspection vehicles. No prior assumptions are made on the 2D location of the traffic signs within each video frame. Rather by sliding a window or cascade detector– of fixed spatial ratio– at multiple scales, candidates for traffic signs are initially extracted from the 2D video frames. Each candidate is then fed into the multi-class traffic sign detection and classification methods. To accurately localize each detection, a non-maxima suppression method is used on the scores of the classifiers to remove those multiple inferences that are caused due to overlapping sliding windows.

The detection and classification methods are as shown in Figure 3.14 are: (1) Haar-like features with Adaboost classifiers; (2) Histogram of Oriented Gradients (HOG) with Support Vector Machine (SVM) classifiers; and (3) a new variant of HOG features where histogram of local color distributions are formed and concatenated with the HOG descriptors to leverage both shape and color information for multiple traffic sign category with one-vs.-all SVM classifiers. Methods (1) and (2) are used in several state-of-the-art methods, and thus our experiments allow the performance of the third method – which is rather new variant of the second method– to be compared and benchmarked against existing methods using linear and non-linear classifiers. In this following, a brief overview of these methods are discussed.

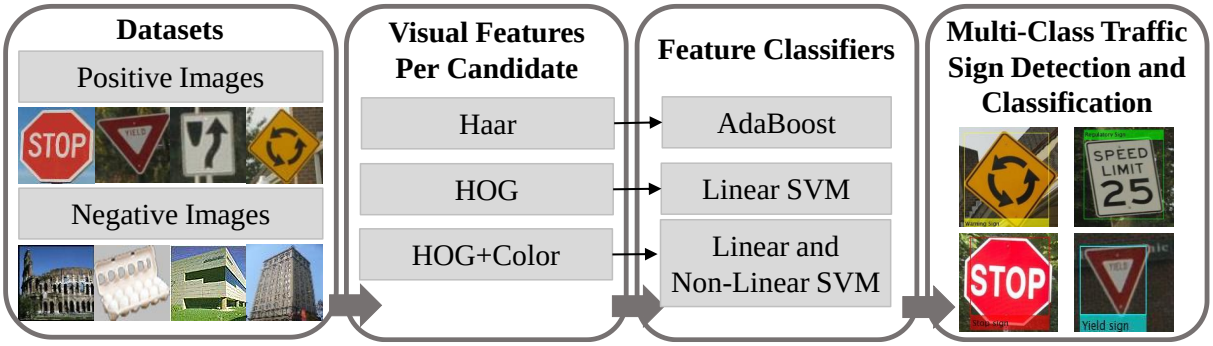

Figure 3.14 Overview of Proposed System per Sliding Window Candidate for Multi-Class Traffic Sign Detection and Classification Using Haar, HOG, and HOG+C Features Together with Adaboost and SVM Classifiers

3.3.1. Method 1-Haar-like feature + Cascade detectors of Adaboost classifiers

The first detector is a cascade of boosted Haar-like classifiers which uses AdaBoost learning method of (Viola and Jones 2001). Here, the initial candidates for traffic signs are convolved using Haar-like features and are ultimately categorized into multiple categories of similar traffic signs (e.g., warning, yield) or simply discarded not containing a traffic sign. The pixel intensities of the adjacent rectangular regions in the convolved images are summed up and the differences are calculated to form the Haar-like features. Each feature is then paired with a threshold and the decision of the classifiers is determined by comparing the feature with the threshold. In this research, we use six different types of Haar-like features which are shown in Figure 3.15. These features can be calculated in real-time, are independent of different image resolutions, and are robust to noise and changes in illumination. They also can be easily scaled to detect traffic signs at various spatial scales that are not presented in the training datasets. For training and testing, separate categories of positive and negative images are put together. Each detector (one per traffic sign category) is trained using distinct sets of positive and negative samples where the positives include the category of interest, while the negatives include other categories as well as the generic set of background images. The overall method is shown in Figure 3.15.

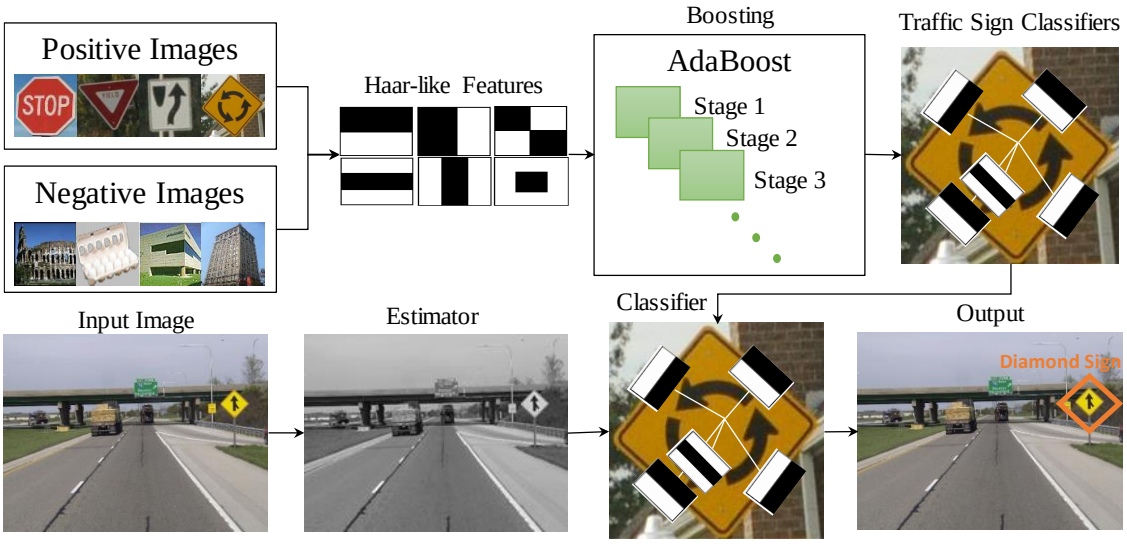

Figure 3.15 An Overview of Training and Testing Process for Haar-like Feature Method

a. Cascade detector for the Haar-like features

The vision cascade object detector extracts the traffic sign candidates from the 2D video frames by sliding a window over all video frame pixels. For each 2D candidate, the detection and

classification method presented above is used to identify if it contains the traffic sign of interest. In this research, we preserve a fixed aspect ratio of 1:1 for the candidates, but we allow the window size to change to detect traffic signs at multiple spatial scales. The cascade detectors are sensitive to out-of-plane rotations and thus may not work well for various traffic signs that have different aspect ratios. To address this limitation, we train a unique detector for each orientation. Table 3.4 shows the parameter values in our cascade detector. These required parameters were analyzed thoroughly to achieve the most efficient performance based on the computational time and accuracy.

Table 3.4 Cascade Detector Parameters

Parameter	Value
False Alarm Rate	0.2
Number of Cascade Stages	5
True Positive Rate	0.995

In Table 3.4, the false alarm rate is the fraction of the negative training samples that are incorrectly classified as positive samples. Increasing the number of stages may result in a more accurate detection result but it will also increase the training time and will requires larger training datasets. The true positive (TP) rate is the fraction of correctly classified positive training samples to all samples.

b. Adaboost classifiers for Cascade detectors

AdaBoost is a technique for combining a number of weak classifiers into a strong one. The results in (Brkic 2013) shows that the method converges to the optimal solution with a sufficient number of weak classifiers. To do so, the AdaBoost assigns weights to weak classifiers based on their quality and the resulting strong classifier is a linear combination of weak classifiers with the appropriate weights. Each stage of the classifier labels the region defined by the current location of the sliding window detector as either positive or negative. If the label is negative, the classification of this region is complete, and the detector slides the window to the next location. If the label is positive, the classifier passes the region to the next stage. The detector reports an object found at the current window location when the final stage classifies the region as positive.

These steps are designed to quickly reject negative samples as the working assumption is that vast majority of the candidate windows do not contain traffic signs. Conversely, TPs are rare,

and worth taking the time to verify their presence. To work well, each stage in the cascade must have a low FN rate. If a stage incorrectly labels a traffic sign as negative, the classification stops, and it would not be possible to correct the misclassification. However, each stage may have a high FP rate. Even if it incorrectly labels a non-traffic sign as positive, the misclassification can be corrected by subsequent stages. The overall FP rate of the cascade classifier is (f^s), where f is the FP rate per stage in the range [0 1], and s is the number of stages. Similarly, the overall TP rate is (t^s), where t is the TP rate per stage in the range [0 1]. Thus, adding more stages reduces the overall FP rate, but it also reduces the overall TP rate. Figure 3.16 shows the process of cascade detector training.

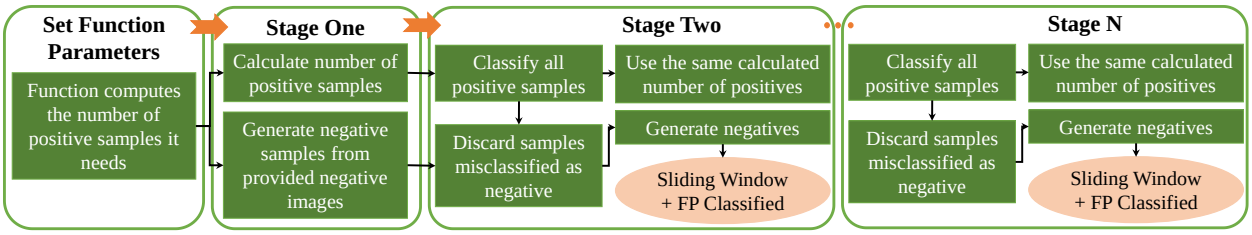

Figure 3.16 Training Process of the Cascade Traffic Sign Detectors

3.3.2. Method 2-Histogram of Oriented Gradients (HOG) with linear SVM classifiers

This method creates histograms of gradient orientations on patches of the images (sliding window candidate) and then compares them to known histograms for specific traffic signs as templates. The basic idea is that the local shape and appearance of traffic signs in a given detection window can be characterized by distribution of the local intensity gradients. These properties can be captured via HOG descriptors of (Dalal and Triggs 2005). In order to do so, the magnitude $|\nabla f(x,y)|$ and orientation $\theta(x,y)$ of the intensity gradient for each pixel with the detection window is calculated. Then the vector of all these orientations and their magnitude is quantized and summarized into a HOG. More precisely, the detection window (Figure 3.17(a)) is divided into $dx \times dy$ local spatial regions (cells) where each cell contains $m \times n$ pixels (Figure 3.17(b)). Each pixel casts a weighted vote for an edge orientation histogram bin, based on the orientation of the image gradient at that pixel. These votes are then accumulated into n evenly-spaced orientation bins over the cells (Figure 3.17(c)). A naïve distribution scheme in form of voting to the nearest orientation bin creates aliasing effects due to under-sampling. Similarly, pixels near the cell boundaries can also produce aliasing along spatial dimensions. The outcome of this process is a

HOG descriptor for each detection window. Similar to (Felzenszwalb et al. 2010), it is possible to use an augmented low-dimensional HOG features leading to a 31-dimensional feature vector. Compare to original 36-dimension features in (Dalal and Triggs 2005), (Felzenszwalb et al. 2010) shows this modification improves the performance. It is hypothesized that the HOG descriptors will be robust enough to intra-class traffic sign variations and lighting changes. This hypothesis is validated in the experimental results section.

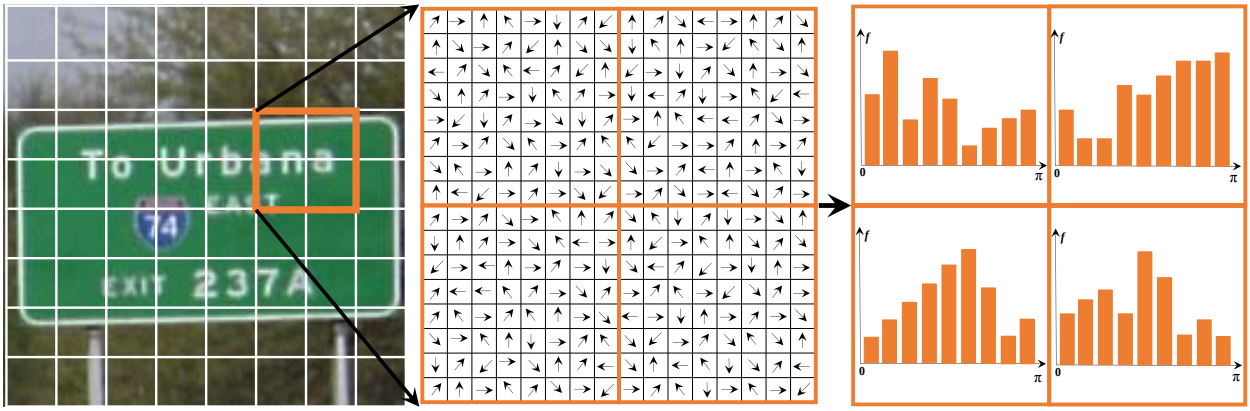

Figure 3.17 Formation the HOG per Sliding Window Candidate: (a) 64×64 Pixel Detection Window, (b) 4×4 Pixel Cell in Each Window, and (c) HOG Corresponding to 4 Cells

a. Linear Support Vector Machine (SVM) classifier

An SVM classifies data by finding the best hyper-plane that separates all data points of one class from those of the other class. The best hyper-plane for an SVM means the one with the largest margin between the two classes. Margin means the maximal width of the slab parallel to the hyper-plane that has no interior data points. The support vectors are the data points that are closest to the separating hyper-plane; these points are on the boundary of the slab.

To train the template models of the traffic signs for the task of detection and classification, a multi-class one-against-all Support Vector Machine (SVM) classifier is employed here. The SVM is a discriminative machine learning algorithm which is based on the structural risk minimization induction principle. Here, the SVM machine learning discriminative classifier is used to identify whether or not the detection window contain a given category of traffic sign. Multiple independent one-against-all SVMs classification approach are developed which each SVM is one of the margin-based classifiers (Burges 1998) and can recognize a specific category of traffic signs in 2D sliding window candidate. As with any supervised learning mode, first the support vector machines are trained and then the classifiers are cross validated. The trained models

are then used to classify (predict/infer) the label of the new observations. Because our training dataset contains considerable number of traffic sign examples, hence we assume that the training data can be linearly separated using linear kernels and as a results the classification can be formulated as follows.

Given *n* labeled training data points $\{x_i, y_i\}$, wherein $x_i\,(i = 1,2,....,n, x_i \in R^d)$ is the set of *d*-dimensional HOG descriptors calculated from each sliding window candidate *(i)*, and $y_i \in \{0,1\}$ is the binary label of a given traffic sign (e.g., stop sign or non-stop sign), the SVM classifier derives an optimal hyper-plane $w^T x + b = 0$ between the positive and negative samples. It is assumed that there is no prior knowledge about the distribution of the resource class video frames. Hence the optimal hyper-plane is the one which maximizes the geometric margin (γ) as shown in Equation (4.11):

$$\gamma = \frac{2}{\|w\|} \tag{4.11}$$

The presence of noise and occlusions which is typical in roadway data collection video streams produces outliers in the SVM classifiers. Hence the slack variables ζ_i are introduced and consequently the SVM optimization problem can be written as:

$$\begin{aligned} &\min_{w,b} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{N}\xi_i \\ &subject\ to: y_i(wx_i + b) \geq 1 - \xi_i \qquad for\ \ i = 1,2,...,N \\ &\xi_i \geq 0 \qquad\qquad\qquad for\ i = 1,2,...,N \end{aligned} \tag{4.12}$$

Where *C* represents a penalty constant which is determined by a cross validation technique. The inputs to the learning (training) algorithm are the training examples for different types of traffic signs and the outputs are the trained models for detection of various traffic signs.

To effectively classify the testing candidates with the HOG descriptors, we slide the detection windows over each video frame at multiple spatial scales with a fixed aspect ratio. In this research, comparison is accomplished by rescaling each sliding window candidate and transforming the candidates to the spatial scale of each template traffic sign model. For detecting and classifying multiple categories of traffic signs, we leverage multiple independent one-against-

all classifiers, each trained to detect one category of traffic signs. Once these models are learned in the training process, the candidate windows are placed into these classifiers and one label from the classifier with the maximum classification score is returned.

b. Localizing traffic signs in 2D

The proposed method for detecting and classifying traffic signs from the entirety of the 2D frames involves application of a detection sliding window. The basic idea is that the detection window scans across each video frame by observing most video frame pixels. At each location, several spatial scales are used for the sliding window candidate to account for scale variations. As shown in Figure 3.18 during this process, the sliding detector window is tiled with a grid of overlapping blocks in which the HOG features will be extracted. The detector window is analyzed and classified whether it contains a particular type of traffic sign or not. This strategy provides a key benefit of detection of traffic signs in close proximity of each other in the video frame under high degrees of occlusions.

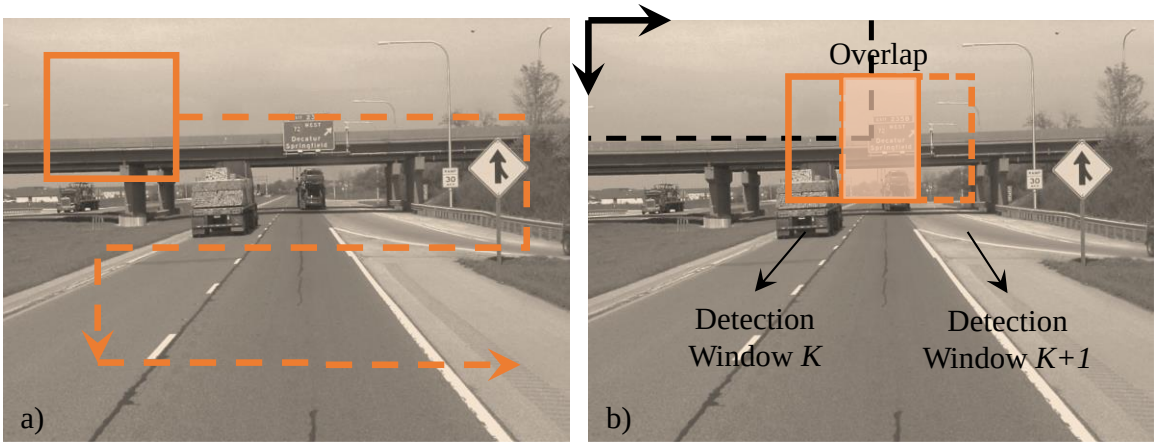

Figure 3.18 Representation of Sliding Window and Extraction of Candidates from the Video Frames

c. Non-maxima suppression for accurate 2D localization

Final step is to find the best detections in each window by selecting the strongest responses within a neighborhood within an image and across scales. The non-maximum suppression module sets all pixels in the current neighborhood window that are lower than the maximum value in that window to zero. This involves grouping the "raw" bounding boxes into equivalence classes based on closeness, deleting groups that contain fewer than a threshold number of boxes, and removing

any groups that lie too close within another group. This has the effect of suppressing all image information that is not part of local maxima and removes overlapping detections with lower scores.

3.3.3. Method 3-Histogram of Oriented Gradients + Color with SVM classifiers

In the last method, we augment the performance of the HOG features with the multiple one-vs.-all SVM classifiers (linear and non-linear) – which are expected to perform best for recognizing the shapes of the traffic signs– with color information. Similar to the HOG features, we divide each sliding window candidate into 8×8 non-overlapping pixel region known as cells. A similar procedure is followed up to compute color attributes for each cell, resulting a histogram representation of local color distributions. Simultaneous to the formation of the HOG descriptor, a histogram of colors is also generated. In order to keep invariance to illumination changes, instead of using RGB color space, HSV color space is used (Gomez-Moreno et al. 2010). To minimize the impact of image brightness, we only use hue and saturation color channels. The local distribution of color is then represented with a histogram that counts the occurrences of a set of evenly spaced normalized hue and saturation values. The color space is then vector-quantized into a 6 bins for hue and 6 bins for saturation to generate color descriptors. These 11-dimensional color descriptors are locally concatenated with the 31-dimensional HOG to form HOG+C descriptors per sliding window candidate. Figure 3.19 summarizes the process.

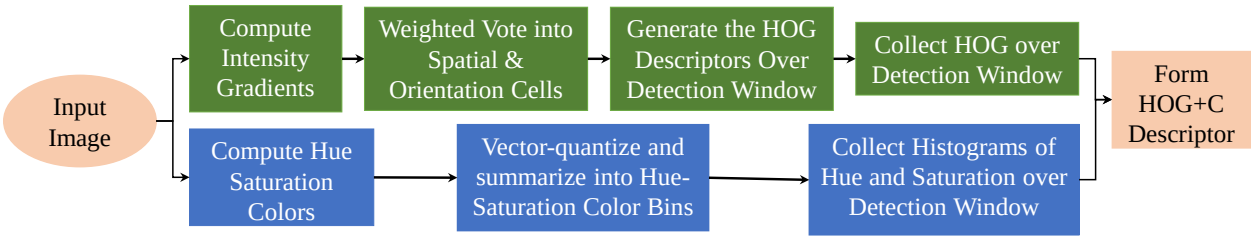

Figure 3.19 Overview of Proposed Method for Forming HOG+C Descriptors

a. Non-linear Support Vector Machine (SVM) classifier

Some binary classification problems do not have a simple hyper-plane as a useful separating criterion. There is a variant of the mathematical approach for these problems that retains the simplicity of an SVM separating hyper-plane. This approach uses these results from the theory of reproducing kernels. The mathematical approach using kernels relies on the computational method of hyper-planes. All the calculations for hyper-plane classification use nothing more than dot products. Therefore, nonlinear kernels can use identical calculations and solution algorithms,

and obtain classifiers that are nonlinear. The resulting classifiers are hyper-surfaces in some space *S*, but the space *S* does not have to be identified or examined. There is a class of functions *K(x,y)* with the property as shown in Equation (4.13). There is a linear space *S* and a function φ mapping *x* to *S* such that

$$K(x, y) = <\varphi(x), \varphi(y)> \tag{4.13}$$

This class of functions includes:

- **Polynomials:** For some positive integer *d*,

$$K(x, y) = (1 + < x, y >)^d \tag{4.14}$$

- **Radial Basis Function (RBF Gaussian):** For some positive number σ,

$$K(x, y) = \exp(-<(x - y), (x - y)> /(2\sigma^2)) \tag{4.15}$$

Here, we also test the performance of these non-linear classifiers against the linear hyper-plane.

3.4. Mapping Traffic Signs Using Google Street View Images for Roadway Inventory Management

This section presents a new system for creating and mapping comprehensive inventories of traffic signs using Google Street View images. As shown in Figure 3.20, the system does not require additional field data collection beyond the availability of Google Street View images. Rather by processing images extracted using Google Street View API using a computer vision method explained in previous section, traffic signs are detected and categorized into four categories of regulatory, warning, stop, and yield signs. The most probable 3D location of each detected traffic signs is also visualized using heat maps on Google Earth. Several data mining interfaces are also provided that allow for better management of the traffic sign inventories. The key components of the system are presented in the following.

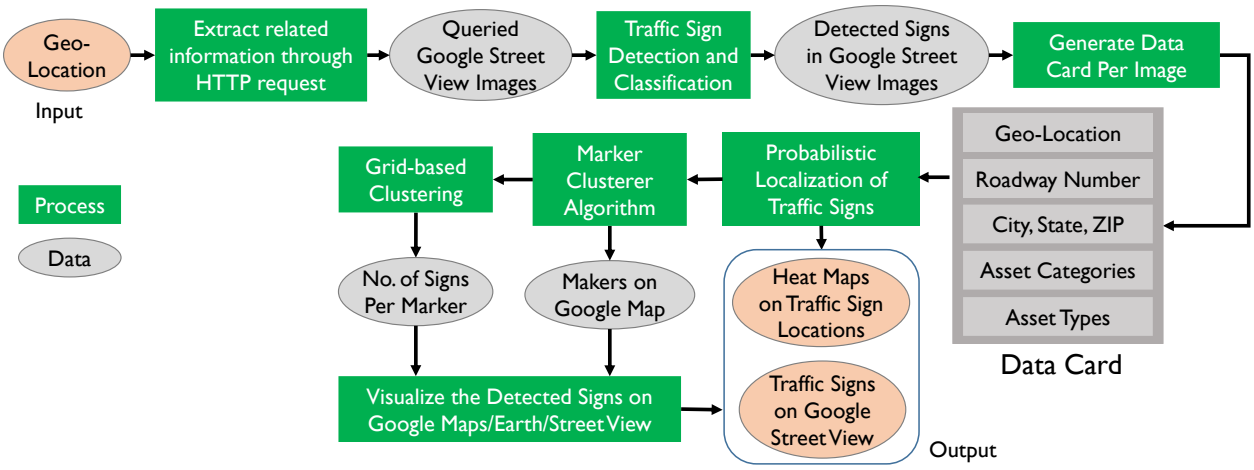

Figure 3.20 Overview of the Data and Process

3.4.1. Extracting Location Information using Google Street View API

To detect and classify traffic signs for a region of interest, it is important to extract street view images from a driver's perspective so that the traffic signs can exhibit maximum visibility. To do so, the user of the system provides latitude and longitude information. The system takes this information as input and through an HTTP request, Google Street View image are queried at the spatial frequency of one image per 10 meter via Google Maps static API. Since the exact geo-spatial coordinates of the street view images are unknown, the starting coordinates are incremented in a grid pattern to ensure that the area of interest is fully examined. Once the query is placed, the Google Direction API will return navigation data in JSON file format, which will then be parsed to extract the polylines that represent the motion trajectory of the cars used to take the images. The polylines will be further parsed to extract the coordinates of points that define the polylines along the road of interest. By adding 90 degrees to the azimuth angle between each two adjacent points, the forward-looking direction of the Google vehicle is extracted. The adopted strategy for parsing the polyline enables identification of the moving direction for all straight and curved roads as well as ramps, loops, and roundabouts. These coordinates and direction information are finally fed into the developed API to extract the Street View images at the best locations and orientations. Figure 3.21 shows the Pseudo code for deriving the viewing angles for each set of locations, where *atan2* returns the viewing direction θ at location (x, y) [between $-\pi$ and π].

Input: A set of location points on the roadway (loc)
Output: A set of forward heading angles for each location (H)

for each two consequtive location points (loc_i, loc_{i+1})
  get latitude and longtitude in (loc_i, loc_{i+1})
  $lat_i \leftarrow loc_i.latitude$; $lon_i \leftarrow loc_i.longtitude$;
  $lat_{i+1} \leftarrow loc_{i+1}.latitude$; $lon_{i+1} \leftarrow loc_{i+1}.longtitude$;
  Append
  $atan2(cos(lat_i) \times sin(lat_{i+1}) - sin(lat_i) \times cos(lat_{i+1}) \times cos(lon_{i+1} - lon_i), (sin(lon_{i+1} - lon_i) \times cos(lat_{i+1})))$
  to headings list (H)
end for
return H

Figure 3.21 Algorithm for Extracting Location Information

This API is defined with URL parameters which are listed in Table 3.5. These parameters are sent through a standardized HTTP which links to an embedded static (non-interactive) image within the Google database. While looping through the parameters of interest, the code generates a string matching the HTTP request format of the Google Street View API. After the unique string is created, the *urlretrieve* function is used to download the desired Google Street View images.

Table 3.5 Required Parameters for Google Street View Images API

Parameter	Description	Dimension
Location	Latitude and Longitude	lat/long value
Size	Output size of the image in pixels.	2048×2048
Heading	Compass heading of camera	0-360 (North)
FOV	Horizontal field of view of the image	90 degree
Pitch	up/down angle of the camera relative to the Street View vehicle	0

3.4.2. Detection and Classification Traffic Signs Using Google Street View Images

It is assumed that each sign is visible from a minimum of three views. A sign detection is considered to be successful if detection boxes (from the sliding windows) in three consecutive images have a *min* overlap of 50%. This constraint is enforced by warping the image after and before of each detection using homography transformation (Hartley and Zisserman 2003). For discriminative classification of the detected traffic signs into multiple categories, method explained in Section 4.3.3 is used.

3.4.3. Mining and Spatial Visualization of Traffic Sign Data

The process of extracting traffic sign data including how TP, FP, and FN detections are handled is key to the quality of the developed inventory management system. Because each sign is visible in multiple images, it is expected that the missed traffic signs (FNs) in some of the images will be successfully detected in the other overlapping images. Thus, the developed system significantly lowers the rate of FNs per traffic sign. In the developed visualization, the most probable location of each detection is visualized on Google Map using a heat map. Hence, those locations that are falsely detected as signs (FPs) – which their likelihood of being falsely detected in multiple overlapping images is small- could be easily detected and filtered out. The adopted strategy for dealing with FNs and FPs significantly lowers these rates (the experimental results validate this). In the following, the mechanisms provided to the users for data interaction are presented:

a. Structuring and Mining Comprehensive Databases of Detected Traffic Signs

For structing a comprehensive database and mining the extracted traffic signs data, a fusion table is developed in which the geo-location information –latitude/longitude– of each detected traffic sign along with type, and correspoding image areas in which the signs are detected. Using Google data management toolbox for fusion tables (Gonzalez et al. 2010), a user can mine the structured data on the detected traffic signs. Figure 3.22 presents an example of a query based on two latitude and longitude coordinates wherein the the number of images in which the detected regularity and warning signs are returned and visualized to the user.

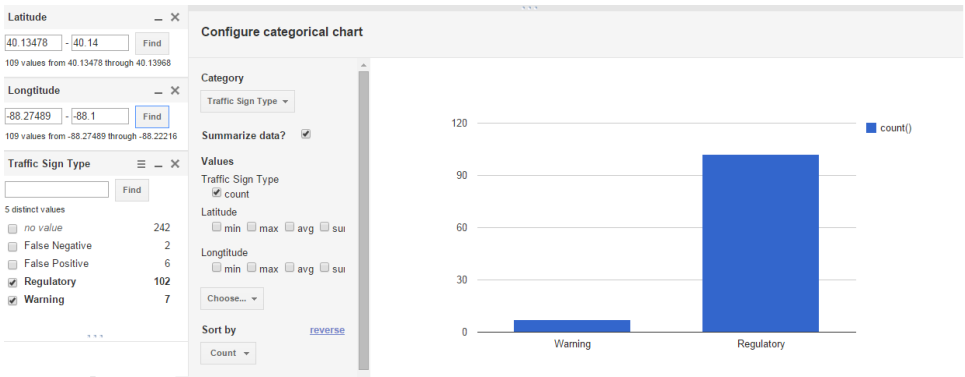

Figure 3.22 Querying the Total Number of Detected Signs and Their Types by Only Specifying Two Latitude and Longitude Coordinates

Figure 3.23 is another example where the analysis is done directly on the spatial data to map detected warning signs between two specified locations.

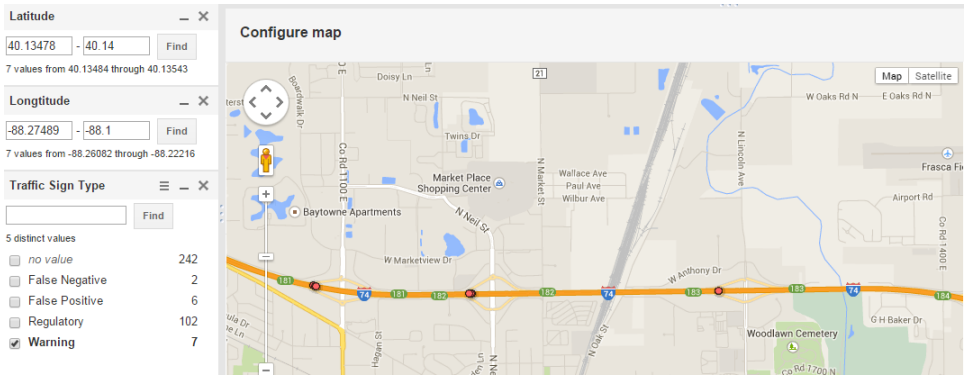

Figure 3.23 Mapping Detected Warning Signs between Two Specified Locations

b. Spatial Visualization of Traffic Signs Data

In the developed web-based platform, Google map interface is used to visualize the spatial data and the relationships between different signs and their characteristics. Google Map, Street View and Earth APIs along with a clustering package and Google fusion table filtering tools were used to develop a dynamic web-based application for visualizing and mining detected traffic signs data. More specifically, a dynamic ASP .NET webpage is developed based on the fusion table that visualizes the result of detected signs on Google Map, Street View, and Earth, by calling needed data using queries from the SQL database and the JSON files.

A javascript is developed to sync a Google map interface with three other views of Google Map, Street View, and Earth (See Figure 3.24). Markers are added for the derived location of each detect sign in this Google Map interface. A user can click on these markers to query the top view (Goole map view), bird-eye view (Google Earth view), and street-level view of the detected sign in the other three frames. In the developed interface, two scenarios can happen:

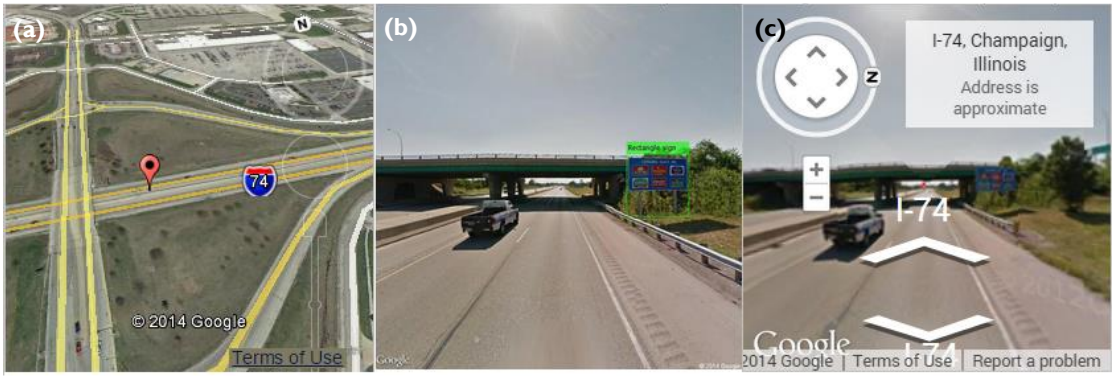

Figure 3.24 Syncing Google Map Interface with Google Earth and Google Street View

Scenario 1. each sign may appear in multiple images– To derive the most probable location for this sign, the area of bounding box in each of these images is calculated. The image that has the highest overall back-projection area is chosen as the most probable 3D location of the traffic sign. This is intuitive, because as the Google vehicle get closer to the sign, the area of the bounding box containing the sign increases.

Scenario 2. Multiple signs can be detected within a single image and thus, a single latitute and longtittue can be assigned to mutiple signs. In these situations, the same as scenario 1 the size of the bounding boxes in images that see these signs is used to identify the most probable location for each of the traffic signs. To show that multiple signs are visible in one image, multiple markers are placed on the Google map.

To visualize these scenarios, the developed interface contains a *static* and a *dynamic* map. In the *static map*, all detections are marked thus multiple markers are placed when several signs are in proximity of one another. Detailed information about latitude/longitude, roadway number, city, state, zip, country, traffic sign type, and likelihood of each detected traffic sign are also shown by clicking on these markers.

To enhance the user experience on the *dynamic map*, the MarkerClusterer algorithm (Svennerberg 2010) is used following by a grid-based clustering procedure to dynamically change the collection of markers based on their distance on the map (depending on the level of zoom). This technique iterates through the markers and adds each marker to its nearest cluster based on a predefined threshold which is the cluster grid size, in pixel. The final result is an interactive map in which the number of detected signs and the exact location of each sign are visualized. As shown in Figure 3.25, a user click on each cluster brings the view closer to smaller clusters until the underlying individual sign markers are reached.

Figure 3.26 shows an example of the dynamic heat maps which visualize the most probable locations for the detected traffic signs. The color coding scheme based on the metaphor of traffic light colors is used in which the colors change as the user zooms in and out of the map. As one get closes to a sign, the most probable location is visualized using a line perpendicular to the road axis. This is because the GPS data cannot differentiate whether a detected sign is on the right side of the road, is on top of a structure in the middle of the view or is on far left. Figure 3.27 presents the Pseudo code for mining and representing traffic sign information.

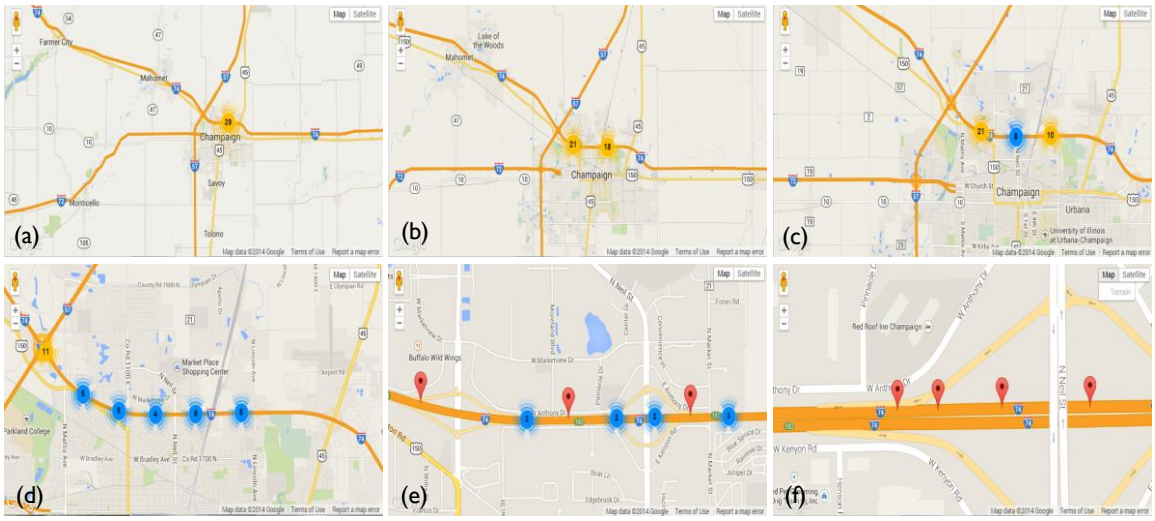

Figure 3.25 The Dynamic Map Interface Wherein by Further Zooming in (or Clicking on the Markers), the More Exact Location of Each Image Containing a Traffic Sign Is Shown. The Numbers Shown Next to the Marker Indicate the Number of Detected Sign in That Section of the Road

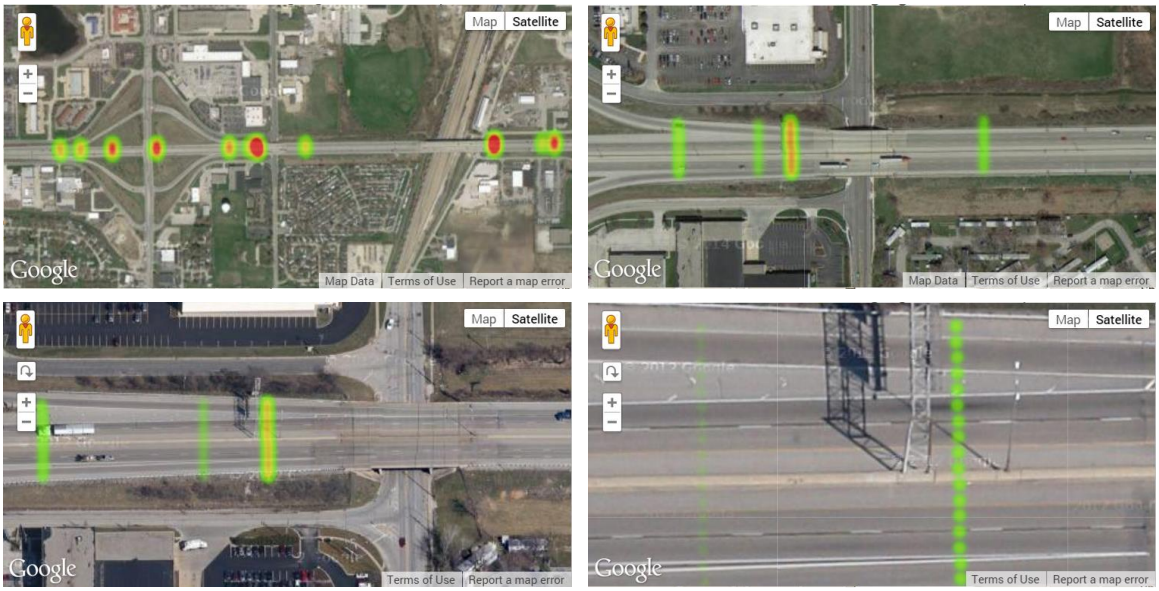

Figure 3.26 Dynamic Heat Map Which Shows the Closest Location of Traffic Signs

Input: Starting and ending point
Output: A database that includes more probable locations and types for all detected traffic signs

```
Set the direction between starting and ending point using
        Google direction API
for each point on the direction
    Find latitude and longtitude of the point
    Calculate FOV (Field of View)
    Get the street view image using the API
    run sign detection procedure
        if traffic sign detected in the picture
            Append to the database the likelihood values for location and type of detected traffic signs
        end if
end for
return Database
```

Figure 3.27 Process of Detecting and Mapping Traffic Signs into the Database

3.5. Image-based Retro-Reflectivity Measurement of Traffic Signs in Day Time

We present a new image-based method to measure retro-reflectivity from a distance during daytime. Our method requires a digital camera equipped with a flashing device. By capturing two images almost simultaneously, the method simulates nighttime visibility and performs retro-reflectivity measurement. Our accuracy and granularity comply with FHWA regulations. The overview of our method is illustrated in Figure 3.28.

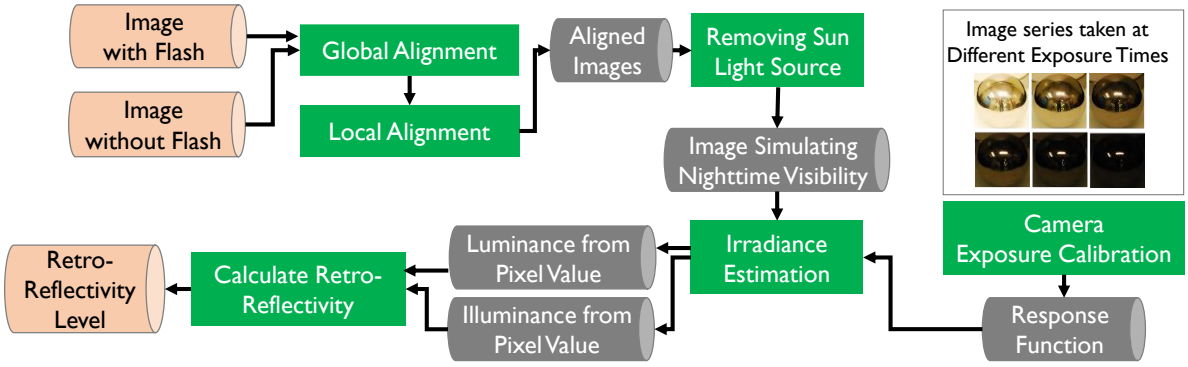

Figure 3.28 Method Overview for Image-based Retro-reflectivity Measurement during Daytime

We use a combination of computational photography and carefully tuned hardware to generate realistic photos of night during daytime. We first capture two images from a scene; one with a controlled artificial light source – produced by a commercial flash, and one without. We

then process the two images to remove all the light sources from the scene except for the controlled light source. Since all natural light sources including the sun are removed, the output image resembles a night photo where only the controlled light source is present. Figure 3.29(a) and Figure 3.29(b) illustrate two photographs taken from the same scene during daytime, one with flash and one without flash. Figure 3.29(c) shows the processed night view of the scene generated by removing natural light sources.

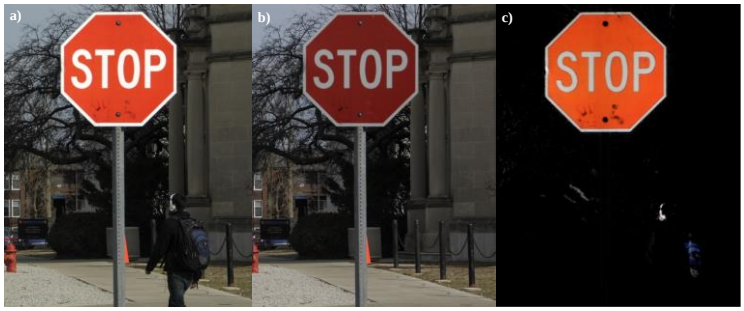

Figure 3.29 (a) Image with Flash; (b) Image without Flash; (c) Our Night Photo Reconstruction

To synthesize a night photo; one needs to remove the sun and all of its reflections while adding a *controlled* light source. To do this, we capture two images where a strong light source is present in only one of the two images. We refer to the image without the artificial light source as I_{Day}. We also refer to the image with the artificial light source present as $I_{Flash+Day}$. We process the two images to estimate the reflection from the controlled light source only. We refer to this image as I_{Flash}. The general strategy to isolate the controlled light source is to first extract true irradiance and then subtract the two images as: $I_{Flash} = I_{Flash+Day} - I_{Day}$. However, a number of important software and hardware measures must be taken before this step is possible. These measures are as follows:

3.5.1. Camera Exposure Calibration

Nearly all cameras apply a non-linear function to recorded raw pixel values in order to better simulate human vision. In other words a pixel *intensity* does not linearly correspond to a pixel *irradiance* (the light incoming to the camera). Therefore direct subtraction would not preserve intensity ratios due to the non-linearity. Rather, this response function depends on camera CCD (Charge Coupled Device) and a number of other factors including exposure time (shutter speed) and f-stop (aperture). Camera is not a photometer and exhibits a limited dynamic range,

with an unknown non-linear response. Hence, in order to convert pixel values to true radiance values, we need to estimate this film response function. The solution is to recover response curve from multiple exposures and then reconstruct the radiance map. For a given camera and with given settings, it is enough to estimate the film response function only once. Then, this response function can be used consistently without the need for recalibrations.

We estimate this response function according to a technique developed by (Debevec and Malik 2008) using multiple images. The goal is to create high dynamic range (HDR) images from low dynamic range (LDR) images and create a HDR tone-mapping. HDR photography is the method of capturing photographs containing a greater dynamic range than what normal photographs contain (i.e. they store pixel values outside of the standard LDR range of 0-255 and contain higher precision). Typically the response function is difficult to estimate. Having multiple observations at each pixel at different exposures, we map image intensities onto a linear space in order to accurately estimate intensity differences. Givens are pixel values Z_{ij} for image with shutter time Δt_j(i^{th}pixel location, and j^{th}image). Exposure is irradiance integrated over time as expressed as Equations (4.16 – 4.18). Then pixel values are non-linearly mapped and rewritten to form a linear system as shown in Equations (4.19 – 4.22).

$$Pixel\ Value\ Z = f(Exposure) \tag{4.16}$$

$$Exposure = Radiance \times \Delta t \tag{4.17}$$

$$\log(Exposure) = \log(radiance) + \log(\Delta t) \tag{4.18}$$

$$E_{ij} = R_i . \Delta t_j \tag{4.19}$$

$$Z_{ij} = f\left(E_{ij}\right) = f(R_i . \Delta t_j) \tag{4.20}$$

$$\ln f\left(Z_{ij}\right) = \ln(R_i) + \ln(\Delta t_j) \tag{4.21}$$

$$g\left(Z_{ij}\right) = \ln(R_i) + \ln(\Delta t_j) \tag{4.22}$$

Given pixel values Z at varying exposures t, the goal is to solve for $g(Z) = \ln(E \times t) = \ln(E) + \ln(t)$ where Z_{ij} gives pixels near 0 or 255 less weight, R_i is radiance at particular pixel site which is the same for each image, Δt_j is known shutter time for image j, and $g(Z_{ij})$ is exposure as a function of the pixel value.

This boils down to solving for $E(irradiance)$ since all other variables are known. By these definitions, g is the inverse, log response function. We use this technique to compute E_{day} and $E_{flash+day}$. This technique is frequently used in High Dynamic Range photography as i is crucial to have comparable pixel intensity values. In this research, we use this technique to estimate the true irradiance.

3.5.2. Automatic Alignment

In order to subtract the two images, I_{Day} and $I_{Flash+Day}$, all elements in the images must be perfectly registered (aligned). A bright edge that is misaligned by one tenth of a pixel (or a few arc-seconds) produces a significant artifact around the edges of objects as shown in Figure 3.30. In real-world conditions, both the camera and objects can move within several pixels. For example, a camera that is fixed on a tripod may be affected by some degree of vibration due to wind and other factors. Furthermore, some objects such as moving cars and pedestrians may move a few pixels between the times that the two images are being captured. A camera that is mounted on a vehicle or a camera that moves could exhibit a greater degree of such misalignments.

Figure 3.30 Misalignment Produces a Significant Artifact around the Edges of Objects

To minimize the effects of vibrations and movements, we automatically align the images in two steps: (1) Global alignment; (2) Local sub-pixel alignment. In global alignment, we translate the images so that the misalignment is below one pixel. In the second step we perform sub-pixel alignment for local areas. The second sub-pixel alignment is essential because at that level the misalignment would be different in various locations of the image.

a. Global Alignment

We compute cross-correlation on the edges of I_{Day} and $I_{Flash+Day}$, to estimate global displacement. Here image edges are used rather than raw pixel intensities because the alignment

of edges is more accurate than the alignment of raw pixel intensities. To speed up the process of cross-correlation, we use Fast Fourier Transform to complete the process in $O(n \log n)$. We then choose the displacement that obtains the maximum alignment (See Figure 3.31). If the camera has not moved, the displacement would be zero. This step makes the algorithm robust to displacements.

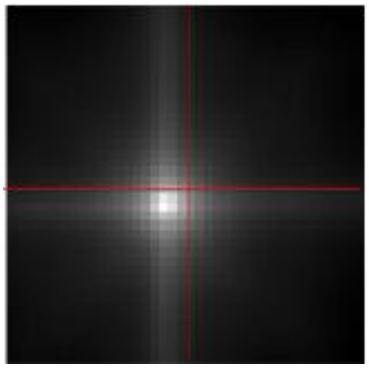

Figure 3.31 Cross-correlation Between the Edges of I_{Day} and $I_{Flash+Day}$

b. Local Alignment

After the process of global alignment some local misalignments may still exist. Misalignment could be different in different areas of the image due to tiny movements in the scene or the camera. This misalignment produces artifacts at the edges. We perform a local sub-pixel registration in order to align every local patch in the image.

The usual technique is to up-sample images and perform a pixel-wise cross-correlation. We use a technique proposed by (Guizar-Sicairos et al. 2008) to align the images without up-sampling. This algorithm registers a pair of images by retrieving the phase difference within a discrete cosine transform. We perform this alignment for a 100×100 grid of patches in the image. Each of the 10,000 blocks are aligned separately according to the local appearance of the block. Figure 3.32 compares the effect of using local alignment on the quality of the output.

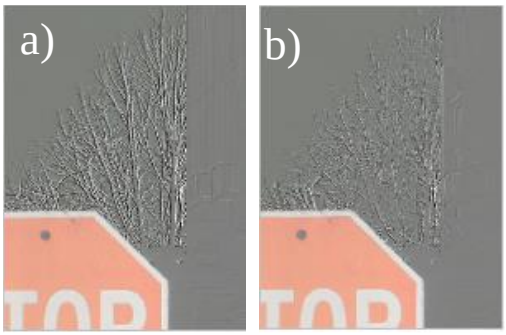

Figure 3.32 (a) Before Sub-pixel Alignment; (b) After Sub-pixel Alignment

c. Irradiance Estimation

After I_{Day} and $I_{Flash+Day}$ are fully registered and the response function g is obtained, we compute E_{Flash} according to the following Equation:

$$E_{Flash} = E_{Flash+Day} - E_{Day} = g(I_{Flash+Day}) - g(I_{Day}) \tag{4.23}$$

3.5.3. Retro-reflectivity Measurement

Retro-reflection is the ratio of the amount of light returned from a traffic sign versus the amount hitting the sign. As shown in Figure 3.33, to determine the retro-reflectivity level of a traffic sign, the measures of luminance, illuminance, and geometry are required.

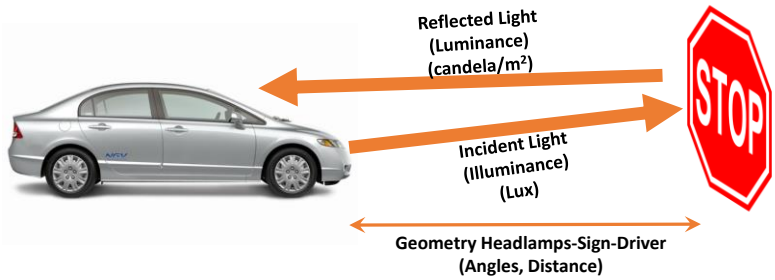

Figure 3.33 Component of Retro-reflectivity

The FHWA has adopted the SI units for retro-reflection; thus by computing the illuminance and luminance, retro-reflectivity is measured in units of candelas per lux per square meter $(cd/lx/m^2)$ as follows:

$$Light\ INTO\ Sign = \ Illuminance = lux\ (lx) \tag{4.24}$$

$$Light\ OUT\ of\ Sign = \ Luminance = \frac{Candela}{m^2}\ (^{cd}/_{m^2}) \tag{4.25}$$

$$Retro - reflectivity = R_A = \frac{Light\ OUT\ of\ Sign}{Light\ INTO\ Sign} \tag{4.26}$$

$$R_A \propto \frac{Luminance}{Illuminance}\ (cd/lx/m^2) \tag{4.27}$$

Equation (4.28) shows the exact relationship between the luminance of a surface (L) in cd/m^2, the illuminance (E) in lx and reflectance (ρ) (dimensionless) where π is the pi number (Hiscocks and Eng 2011).

$$L = \frac{E\rho}{\pi} \tag{4.28}$$

a. Luminance from Pixel Value

The digital camera turns an image into a two dimensional array of pixels. Ignoring the complications of color, each pixel has a value that represents the light intensity. The amount of exposure (the brightness in the final image) is proportional to the number of electrons that are released by the photons of light impinging on the sensor. Consequently, it is proportional to the illuminance (in lux) times the exposure time, so the brightness is in lux-seconds. Invoking the parameters of the camera, we have formula form (Conrad 1998):

$$N_d = K_c \left(\frac{tS}{f_s^2}\right) L_s \tag{4.29}$$

Where N_d is value of the pixel in the image, K_c is calibration constant for the camera, t is exposure time in seconds, f_s is aperture number (f-stop), S is ISO sensitivity of the film, and L_s is luminance of the scene in *candela/meter*^{2}. One measurement would be sufficient to determine the value of calibration constant. One would photograph some source of known luminance L_s, determine the value N_d of the pixels in the image, and note the camera settings for ISO exposure time and aperture. To calculate the N_d, we simply take the red, green, and blue values and use the following formula to convert them into a grayscale pixel:

$$N_d = R \times 0.299 + G \times 0.587 + B \times 0.114 \tag{4.30}$$

The reason these values are weighted is because pure red, green, and blue are actually darker/lighter than each other, with green being the darkest and blue the lightest (Johnson 2006).

b. Illuminance from Pixel Value

American Standard Association (ASA) has defined incident-light meters as well as reflected-light meters. The incident-light exposure Equation is

$$\frac{A^2}{t} = \frac{ES_x}{K_c} \tag{4.31}$$

where A is the relative aperture (f-number), t is the exposure time (shutter speed) in seconds, E is the illuminance, S_x is the ASA arithmetic film speed, and K_c is the incident-light meter calibration constant. By placing the luminance and illuminance in Equation (4.28), we measure the sign retro-reflectivity from every pixel in the images taken during the daytime. Finally, the retro-reflectivity of a sign is measured by averaging all pixel-level measurements. This is consistent with current practices, yet instead of using a few (typically up to 4) point-level measurements, all pixels are used to characterize retro-reflectivity more accurately.

CHAPTER 4. DISCUSSION AND EXPERIMENTAL RESULTS

4.1. Segmentation and Recognition of Roadway Assets using Image-based 3D Point Clouds and Semantic Texton Forests

As an initial step, the performance of the proposed algorithms for the segmentation and recognition of the roadway assets is evaluated at Virginia Tech's Smart Road. Using the Smart Road facility, a new dataset for benchmarking the performance of our 3D reconstruction, 2D segmentation, 3D labeling, and localization algorithms was created. This dataset is for both training and testing purposes so that it can be released to the community for further development and validation of new algorithms. For this purpose, we collected 30 minutes of video streams that were recorded using one single camera pointing towards the assets on the right side of the road.

4.1.1. Data collection and setup

Our entire training and testing dataset includes a total of 270 images in 12 different object categories. For each image in both training and testing datasets, a ground truth image is generated in a supervised fashion. These ground truth images are color-coded based on their categories and labels such as asphalt pavement, guardrail, and traffic signs were placed accordingly. Table 4.1 presents these segmentation categories, the number of images used per category, and finally the specific colors that are assigned to each category for supervised training and automated testing purposes. In our ground truth image dataset, those pixels that are non-relevant to selected object categories were intentionally color-coded in black, further highlighting the void category.

Table 4.1 Semantic Segmentation Asset Categories and Their Corresponding Colors

Category Name	Images (#)	Color	Category Name	Images (#)	Color
Asphalt Pavement	45	(0,128,1)	Grass	27	(129,0,127)
Concrete Pavement	13	(1,0,128)	Soil	42	(254,0,0)
Guardrail	47	(128,0,0)	Sky	42	(0,255,1)
Poles	35	(127,128,0)	Safety Cones	11	(0,0,254)
Signs	16	(128,128,128)	Traffic Signals	7	(254,128,254)
Trees	24	(0,128,129)	Pavement Markings	23	(127,255,254)

Several examples of the training images and their ground truth for each asset category are presented in Figure 4.1. As observed, in the ground truth images, those pixels that are corresponding to our categories are color-coded according to Table 4.1. For example, Figure

4.1(c2) shows how the guardrail has been isolated from asphalt pavement, sky, tree, and soil pixels. The STF model is trained using our training images and their ground-truth which consist 70% of our entire dataset. The rest of the 30% were used for testing purposes.

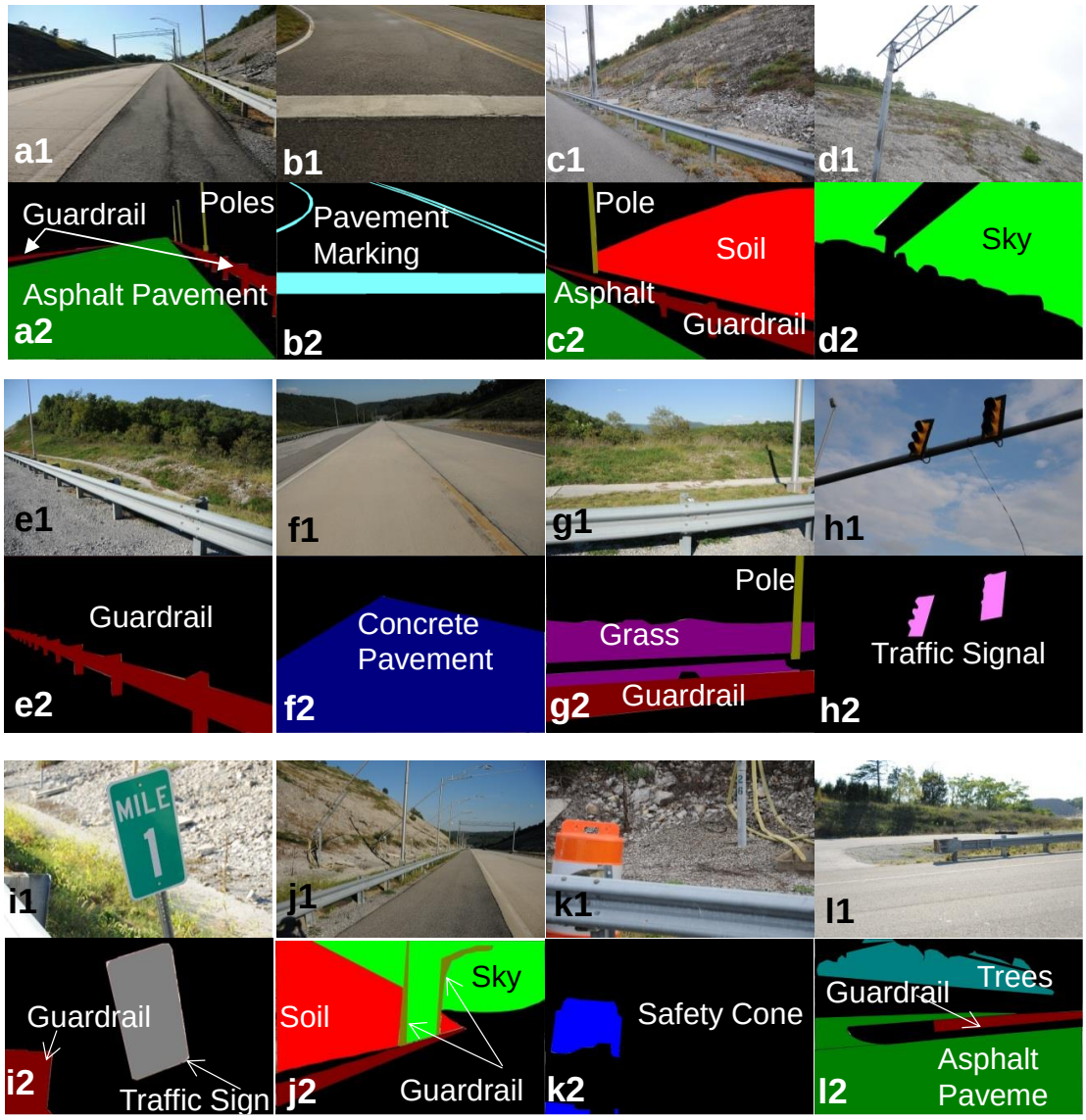

Figure 4.1 Supervised Segmentation of the Ground Truth Images. In Each Block Represented by Alphabetical Letters, the Image Is Shown in "1" and Corresponding Ground Truth Is Shown in "2". The Ground Truth Images Are Color-coded Based Categories Represented in Table 4.1

We used two datasets of 66 and 171 images for reconstruction of the 3D point cloud models. These images represent guardrail, asphalt pavement, soil, and grass categories. Each of these datasets was reconstructed separately and the images that were used for their reconstruction

were color-coded using the STF model. Finally, all the reconstructed points in the 3D point cloud models were color coded based on the outcome of the voting scheme presented in Chapter 3.

4.1.2. Semantic Texton Forest setup

In our experiment, five decision trees were trained on a subset of training asset data and then filled with all of training data points as described in method chapter. We trained a forest using the values for the parameters in Table 4.2. The values are selected based on extensive initial experiments and recommendations from our previous work as well as the original work on STF method (Golparvar-Fard et al. 2012; Johnson and Shotton 2010).

Table 4.2 Parameters Used for Training Scenario

Parameter	Value
Number of Trees	5
Maximum Depth	10
Type of Split Tests	P,P+Q, P log(Q),
q	15
Color Channel	RGB
Data % Per Tree	25%

4.1.3. Evaluation metrics

In our experiments, accuracy of both segmentation and recognition are measured. Particularly the accuracy of segmentation for each type of assets is measured based on the mean percentage of pixels labeled correctly over all asset categories. A confusion matrix was computed over all pixels $p \in P_R$ where P_R is the set of test pixels. Thus, for each pixel $p \in P_R$, using the ground truth images the returned labels are compared with their ground truth *G(p)*, as follows:

$$M[i,j] = \left| \left\{ p : p \in P_R, G(p) = c_i, \arg\max_c P(c \mid L_p) = c_j \right\} \right| \quad (5.1)$$

Next, for all pixels that belong to asset category *i* in an image, (α_i) is calculated according to Equation (5.2). Finally, the mean category accuracy μ_i is reported in the confusion matrix, and is calculated as:

$$\alpha_i = \frac{M[i,j]}{\sum_j M[i,j]} \tag{5.2}$$

$$\mu_i = \frac{1}{Z}\sum_z \alpha_i \tag{5.3}$$

Each index in the confusion matrix (μ_i) shows for each pair of segmented category $<c_1,c_2>$, how many asset categories from c_2 were incorrectly assigned to c_1. Each column of the confusion matrices represents the predicted asset category and each row represents the actual asset category. The segmented True Positives (TP), False Positives (FP), and False Negatives (FN) are compared and the percentages of the correctly predicted categories with respect to the actual category are calculated using the above formulas and represented in each row. A second metric for validation is the overall accuracy α at the pixel level, which we have called it accuracy of recognition in our experiments and is calculated as:

$$\alpha_i = \frac{\sum_i CM[i,j]}{\sum_i \sum_j CM[i,j]} \tag{5.4}$$

The mean category accuracy μ ensures a fair balance across asset categories which potentially have very different numbers of pixels in the data. The overall accuracy α, on the other hand, provides an indication on the proportion of the asset images that can be reliably segmented using our proposed method. Both of these metrics are important to get a sense of the accuracy in the proposed method. For example, a low μ along with a high α can indicate the presence of over-fitting to a particular asset category which is the result of that asset category being disproportionately represented in the dataset.

4.1.4. Experimental results and validation

In this section, the experimental results from the proposed algorithms are presented. The developed segmentation method using STF is implemented in Windows 7 64bit Visual Studio(C#) and is primarily based on the original implementation of (Shotton et al. 2008). The 3D reconstruction pipeline is built in Linux 64bit upon the previous system of (Golparvar-Fard et al. 2012). The experiment was benchmarked on an Intel(R) Core(TM) i7 960 with 24 GBs RAM and

NVIDIA GeForce GTX 400 graphics card. The visualization platform is implemented in C++ using Microsoft DirectX9.0 graphics library.

Figure 4.2 shows several examples of our experimental results on the segmentation of assets. As observed, most parts of these images are properly segmented for the expected assets. Figure 4.3 further presents some examples of those cases where segmentation resulted in wrong categories.

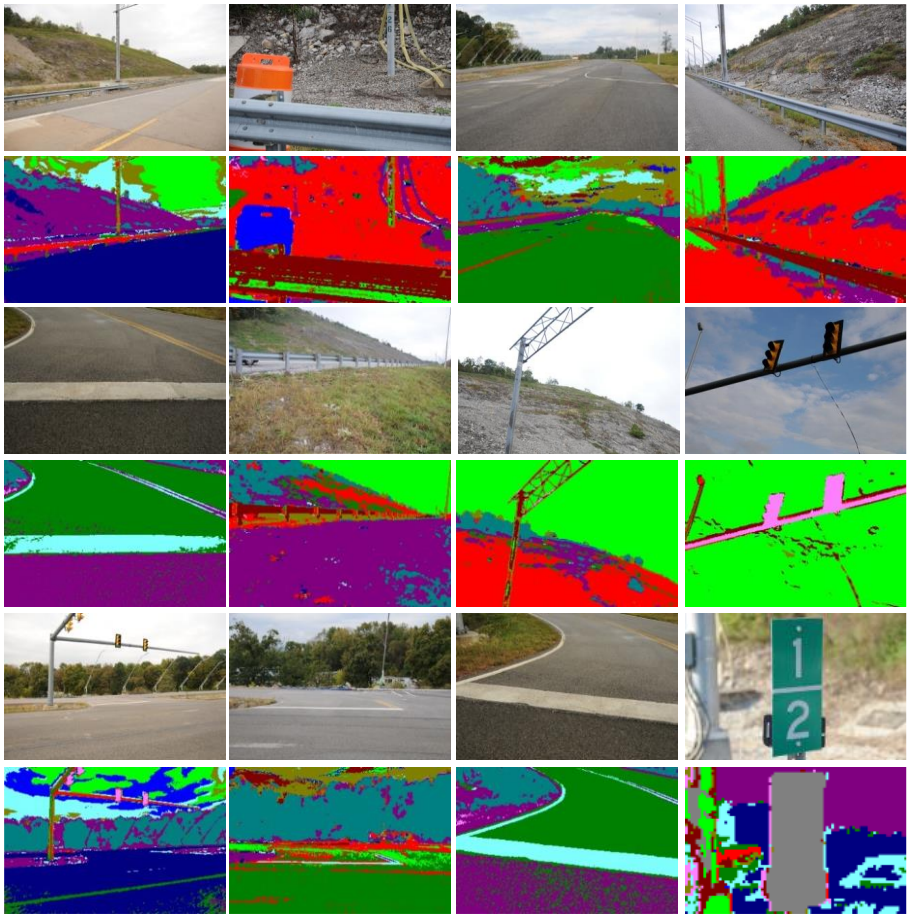

Figure 4.2 Successful Segmentation and Asset Recognition Results; Each Two Rows Show the Original 2D Image and the Outcome of the Segmentation

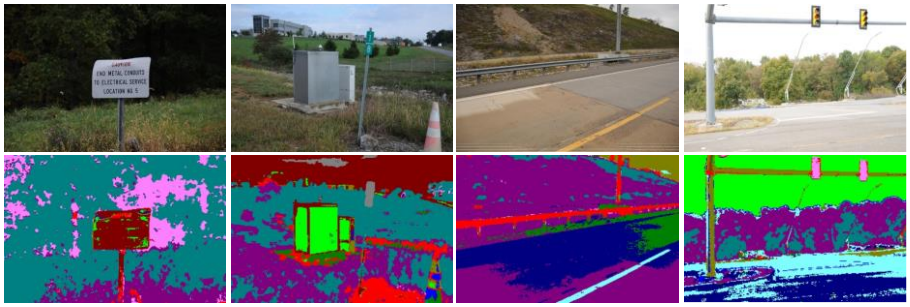

Figure 4.3 False Segmentation and Asset Recognition Results

The resulting 3D point clouds for both experimented datasets are shown in Figures 4.4 and Figure 4.5. Figures 4.6 and Figure 4.7 show the outcome of the labeling for the points in the reconstructed clouds. Such visualization enables the user to select an asset category of interest, and minimizes the search time in finding appropriate imagery. In both of these figures, the outcome is represented in form of D4AR (4 Dimensional Augmented Reality) visualization models, where in the point cloud models and their geo-registered imagery are visualized together. The user can either navigate through the geo-registered imagery, or conduct joint observations to the point cloud and imagery in 3D. The number of images, computational time and the success rate for using imagers in the reconstruction of the 3D point clouds are presented in Table 4.3.

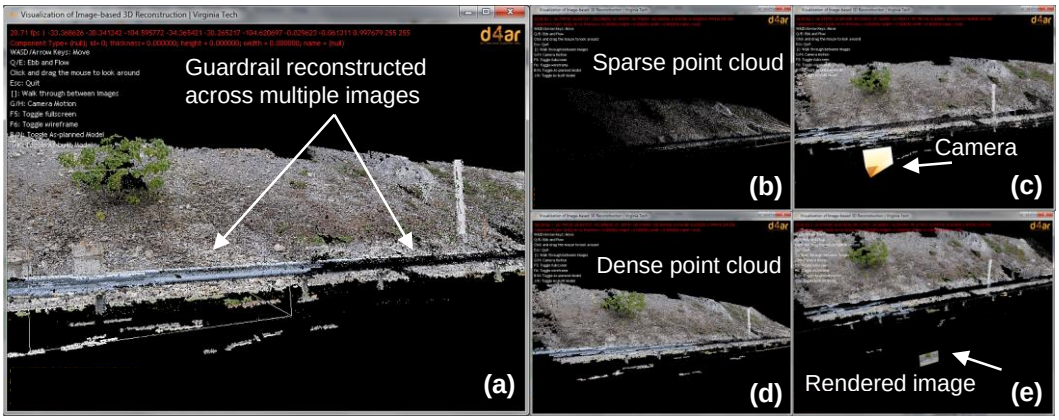

Figure 4.4 3D Image-based Reconstruction Results from Experiment #1

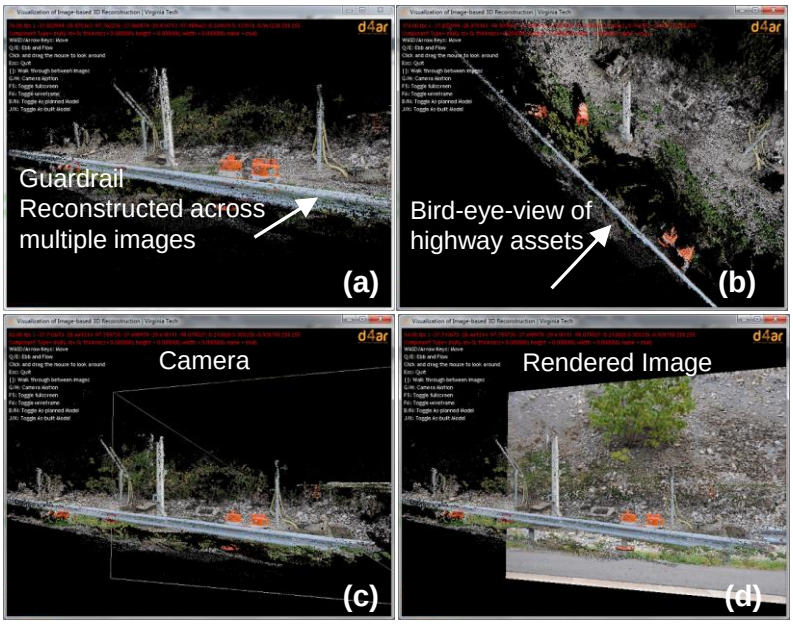

Figure 4.5 3D Image-based Reconstruction Results from Experiment #2

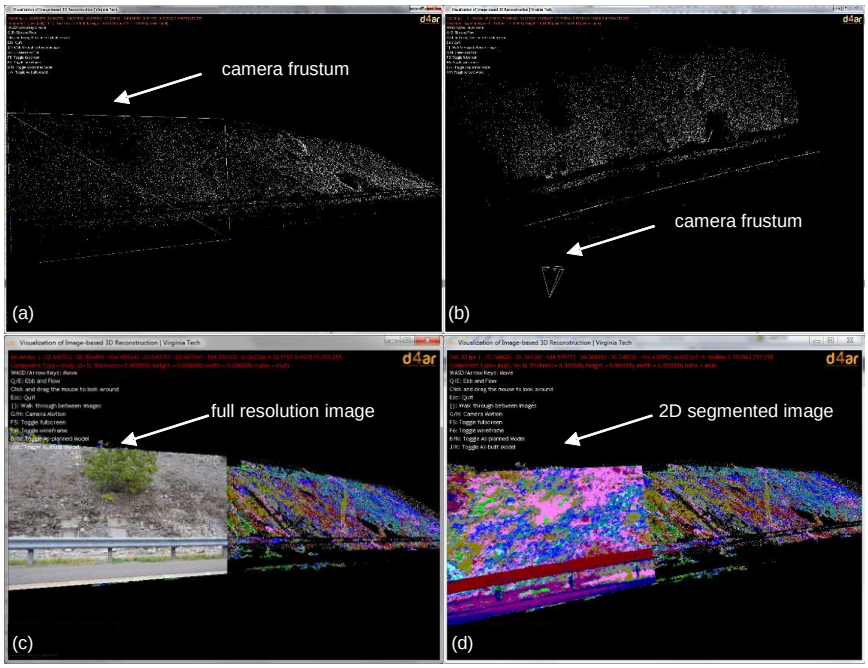

Figure 4.6 3D Image-based Reconstruction Results: (a) Point Cloud Reconstructed Using 66 Images Observed from a Camera Frustum; (b) 3D Location of the Camera; (c) The Camera Frustum Rendered with the Full Resolution Image, and (d) 2D Segmented Image Rendered Over the Camera Frustum

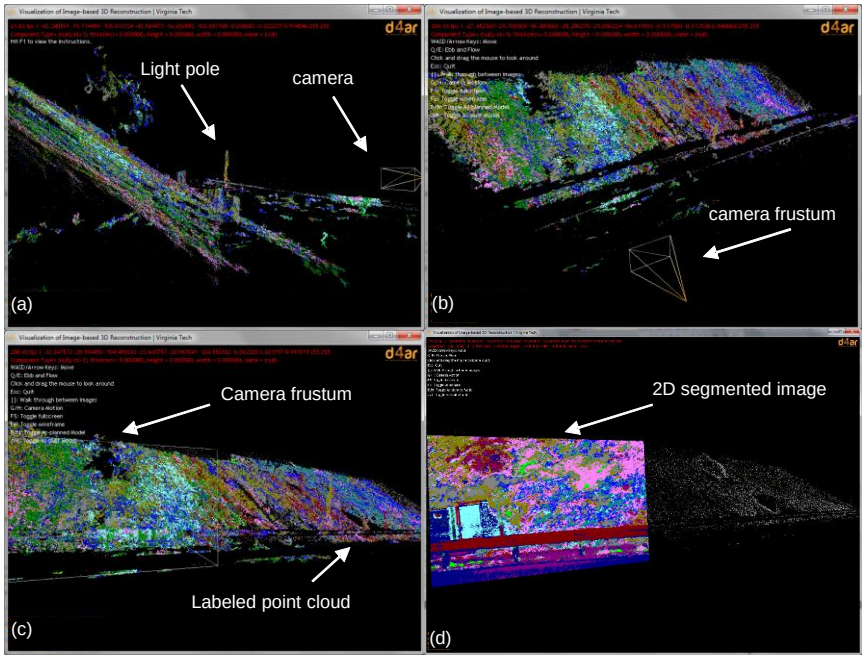

Figure 4.7 3D Image-based Reconstruction Results: (a) Reconstruction of a Light Pole and Correct Labeling. The Location of the Camera to the Road Profile Is Also Shown; (b) The Camera Location with Respect to the Labeled Point Cloud; (c) 2D View to the Point Cloud from a Camera Frustum; and (d) 2D Segmented Image Rendered Over the Camera Frustum

Table 4.3 Result of 3D Image-based Reconstruction

Images (#)	Density*		Computation time			Rate
	SfM	D4AR	SfM	D4AR-v1	D4AR-v2**	
66	663,726	902,355	1hr 21min	3hr 40min	32min	1.00
171	175,737	2,076,887	8hr 54min	10hr 17min	1hr 14min	0.98

* Density in form of the number of points in the reconstructed cloud

** GPU and multi-core CPU-based implementation of the D4AR image-based 3D reconstruction

Table 4.3 highlights the significance of the decrease in computational time compared to our previous work in 3D reconstruction. As observed, the new implementation based on GPU and multi-core CPU has significantly reduced the computation time, making it feasible to reconstruct large areas which are typical in case of roadway infrastructure assets. In the following two sections, the accuracy of asset recognition and segmentation are discussed in more details.

4.1.5. Accuracy of recognition

In order to test the recognition accuracy on both training and testing images, we compare the outcome at the pixel level with their corresponding ground truth images. Hence, the color value of each pixel in the segmented imagery is compared with ground truth. Table 4.4 shows the number of pixels of segmented image with exact color values of those indicated in the ground truth. The average accuracy of recognition for the segmented imagery at the pixel level is 86.75%.

Table 4.4 Accuracy of 2D Image Segmentation

Category	Accuracy of segmentation
Asphalt Pavement	82.58 %
Concrete Pavement	99.04 %
Guardrail	85.81 %
Grass	72.30 %
Traffic Signal	91.78 %
Pavement Marking	89.67 %
Poles	71.77 %
Safety Cones	85.89 %
Traffic Sign	98.05 %
Sky	98.25 %
Soil	87.30 %
Tree	78.62 %

4.1.6. Accuracy of segmentation

Figure 4.8 shows the confusion matrix for segmentation of asset categories. Here, the accuracy refers to the segmentation for each region within the imagery. Overall, an average

accuracy of 76.50% for region segmentation is achieved. Our segmentation works best for traffic signals, safety cones, guardrails and traffic signs. The regions belonging to other objects such as trees, soils, and grass are also properly segmented. As observed, the largest confusion happens between ‘asphalt pavement’ and ‘soil’ categories. Another significant confusion occurs between ‘asphalt pavement’ and ‘concrete pavement’ categories. These are primarily related to the visual consistency of these categories as well as the varying appearance of the soil category.

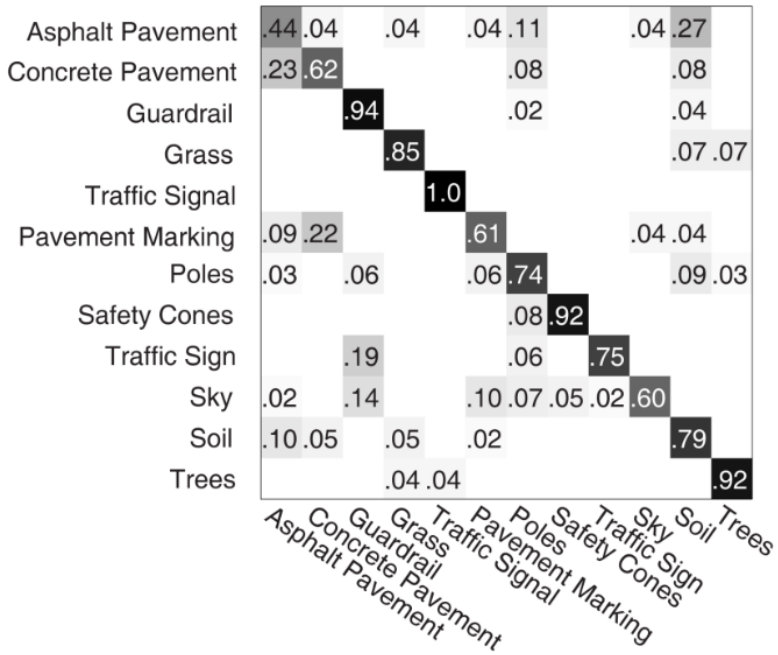

Figure 4.8 Confusion Matrix for 2D Segmentation of Asset Categories

4.1.7. Discussion on the proposed method

Overall, the experimental results are satisfactory given the rather smaller size of the data that has been tested in this study. One observation indicated that having proper images in the training dataset in which categories are visually distinguishable, would help in formation of the STFs and results in better pixel-level recognition and region segmentations. As represented in Figure 4.8 and Table 4.4, the True Positive (TP) rates for the traffic signals are among the highest in our asset categories. In contrary, the segmentation results for the asphalt pavement are among the lowest since the features of these asset items resemble other asset items such as the soil or concrete pavement.

Compared to machine learning algorithms that benefit from filter banks, in our work, the computational time for applying the segmentation is considerably shorter; i.e., is in order of seconds. This further justifies the application of STF algorithm for asset segmentation. Given the

high volumes of roadway assets, minimal computation time is an important attribute for any roadway asset condition assessment system. Furthermore, the combination of the 3D reconstruction and asset categorization enables assets to be localized in 3D; i.e., the user can query <*x*,*y*,*z*> coordinates for any point from 3D. Compared to the application of GPS or wireless which can only represent the existence of a given type of asset in a radius, this work can localize identified assets in a higher precision. The color-coded point cloud models also enable users to easily navigate to areas of interest and conduct condition assessments.

4.2. Segmentation and Recognition of Roadway Assets from Car-Mounted Camera Video Streams using a Scalable Non-Parametric Image Parsing Method

4.2.1. Data collection and setup

We leverage two types of datasets along with their ground truth data for our experiments:

a. Smart Road dataset

The first dataset in our experiments, referred as "Smart Road", comes from (Golparvar-Fard et al. 2012). Smart Road, as shown in Figure 4.9 is a unique, state-of-the-art, full scale, closed test bed 2.2 mile long research facility managed by Virginia Tech Transportation Institute and owned by Virginia Department of Transportation in Blacksburg, VA. Smart Road features a variety of roadway assets and unique capabilities and it is closed to live traffic, which makes it an ideal location for data collection and experiments. We initially videotaped the road assets along the Smart Road to validate several algorithms before conducting our full experimentations. This dataset allowed us to challenge the initial prototyped algorithms and conduct testing on FP and FN predictions.

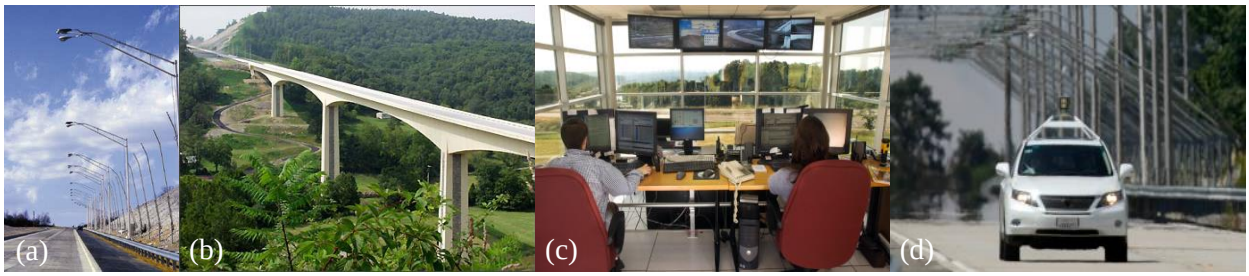

Figure 4.9 Smart Road: (a) Height Adjustable Poles; (b) Virginia's Highest Bridge; (c) Control Room; (d) Google Car Testing on the Smart Road

The training video frames in the "Smart Road" Dataset' contains 200 fully and partially annotated images: 2,169 samples of 12 different classes of roadway assets and 1,119 samples of geometric labels. The frequency histogram of different labels annotated on this dataset are shown in Figure 4.10. Examples of the training images and their ground truth for both semantic and geometric labels on each asset category are presented in Figure 4.11.

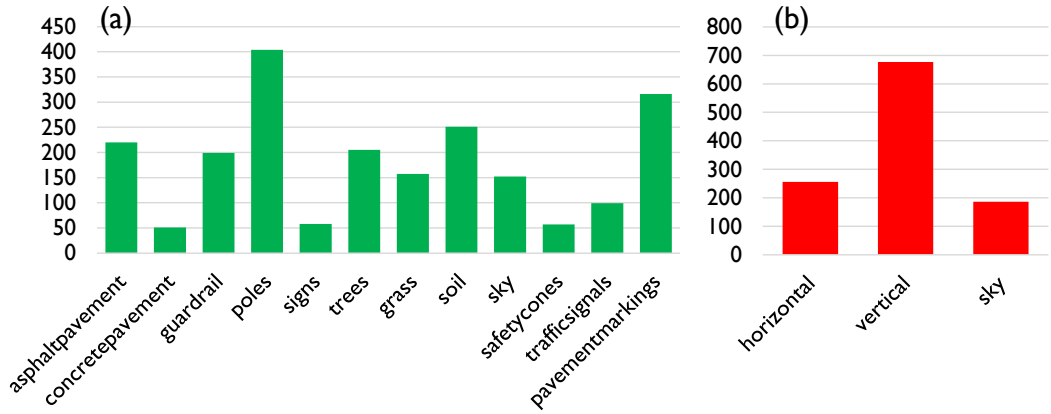

Figure 4.10 Frequency Histograms of Semantic and Geometric Labels on the Smart Road Dataset Assigned to the Superpixels

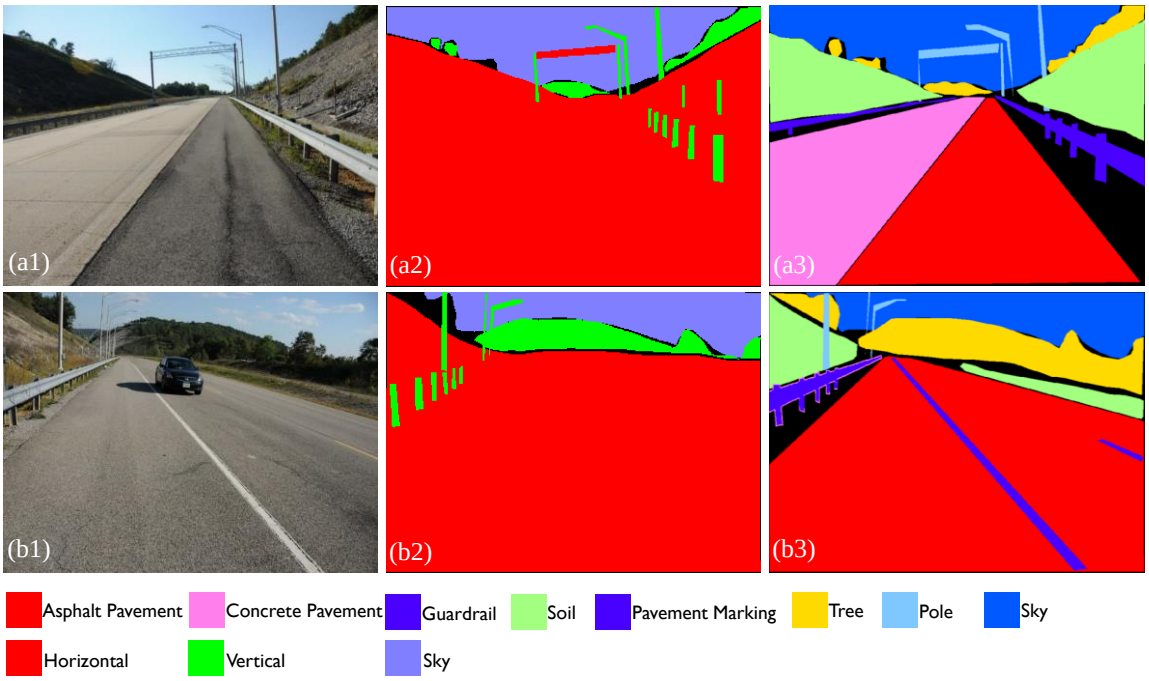

Figure 4.11 Examples of Ground Truth Images for Smart Road Dataset: (1) Actual Image; (2) Geometric Label; (3) Semantic Label

b. Interstate I-57 dataset

The second dataset is collected along the U.S. Interstate 57 by Illinois Department of Transportation. The inspection vehicle is a Ford E350 full size van which can travel and collect images and data at highway speeds. Low speed cut off of sensor data instrument is 15 mph and image collection stops when vehicle stops. There are five cameras including three front view, one rear view, and one down shot for pavement view. The resolution of images is 1300×1000 pixels. The cameras can capture images at a rate of 26.4 feet or 200 images per view per mile. The front view cameras are horizontally located in 45˚ angle with each other and the vertical angle is adjusted in a way to ensure that pavement is primary focus and also can capture overhead signage. All cameras are triggered and a software controls cameras. Figure 4.12 shows the data collection vehicle with mounted cameras.

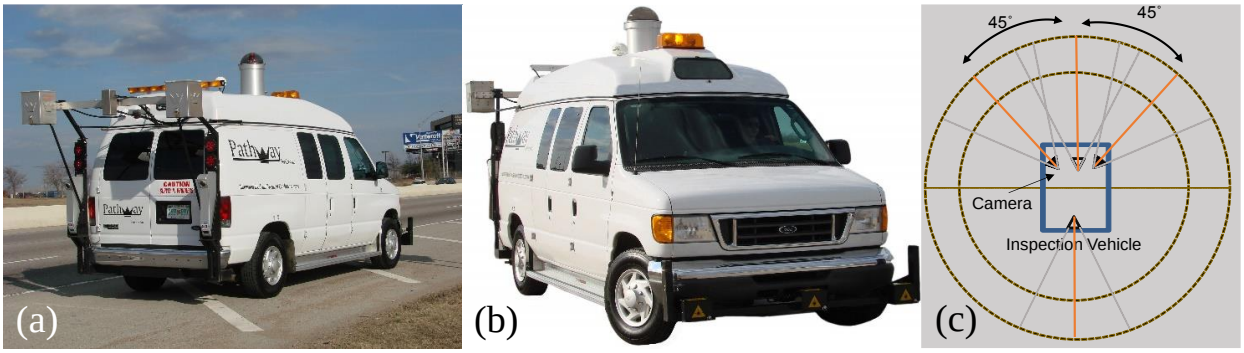

Figure 4.12 Data Collection Using Inspection Vehicle with Three Mounted Frontal-Cameras. This Vehicle Was Used to Collect Imagery Data on I-57 Used for Our Experiments

Using both Smart Road and I-57 highway video frames, a new dataset is created for benchmarking the performance of the proposed method for segmentation and recognition of high-quantity low-cost road way assets. While the training and testing images have only been collected from the same roadways, our method (and the trained model) can be applied to any existing roadway in the U.S. and is scalable because the assets in other roadways will be consistent in form/shape with the ones used for training our system. Our method has also been tested under different viewpoint and lighting configurations. Also our retrieval process provides us with an opportunity to retrieve highway vs. secondary roadway datasets which helps us to narrow down the search on the feature matching process and improves the performance of the overall system. This combined dataset is used for both training and testing and it is released through http://raamac.cee.illinois.edu/aca for future developments and validations.

To generate the ground truth data for our experiments, a subset of training asset data including geo-labels and semantic-labels are generated using LabelMe toolbox which is a web-based image annotation tool (Russell et al. 2008). These annotations are 1) geometric labels, and 2) semantic labels. The geo-labels are color-coded based on categories such as horizontal, vertical, and sky. Similarly, the semantic labels are color-coded based on categories of asphalt pavement, guardrail, safety cone, and traffic signs. In the ground truth video frames, those pixels which were not been labeled, remain as "black". We manually map each semantic category to a unique geometric category. For example, "traffic sign" is "vertical", "guardrail" is "horizontal" and so on. Including the partially annotated images allows our method to manifest if they are able to benefit from additional partially labeled images. There is a dozen of object classes with hundreds of training samples and there are some object classes with just a few of training samples such as safety cone asset category. The main challenges in this training and testing process is that many object classes have very few training samples. For such unbalanced and non-uniform dataset, the performance is evaluated at the per-pixel classification rate.

It contains a training set of 550 images that have been fully and partially labeled by LabelMe toolbox. This dataset has been split into 400 training images along with 5970 labels and 150 test images and used 8 different classes of roadway assets. The frequencies of different classes on this dataset are shown in Figure 4.13. Video frame from categories such as pavement marking, traffic sign, shoulder, and pavement are common, but there are also some categories that rarely appear in roadway datasets like safety cones and traffic signals.

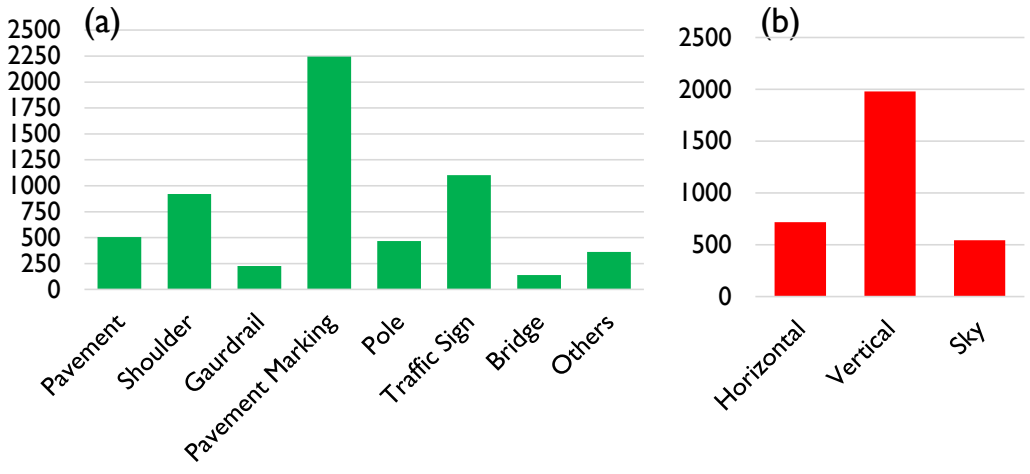

Figure 4.13 Frequency Histograms of Semantic and Geometric Labels on the I-57 Dataset Assigned to the Superpixels

4.2.2. Evaluation metrics

The accuracy of segmentation for each type of asset is measured based on the mean percentage of pixels labeled correctly over all asset categories. A confusion matrix is also computed over all pixels $p \in P_R$ where P_R is the set of test pixels. For each pixel *p*, the returned label are compared with the ground truth $G(p)$ as shown in Equation (5.5):

$$M[i, j] = \left\{ p : p \in P_R, G(p) = a_i, \arg\max_a P(c \middle| L_p) = a_j \right\} \tag{5.5}$$

Then, for all the pixels that belong to asset category *a* in an image, (Ω_a) is calculated according to Equation (5.6). Finally, the mean category accuracy (μ_a) is reported in the confusion matrix.

$$\Omega_a = \frac{M[i, j]}{\sum_j M[i, j]} \tag{5.6}$$

$$\mu_a = \frac{1}{Z} \sum_z \Omega_a \tag{5.7}$$

The second metric for validation the overall accuracy (Ω) at the pixel level – the accuracy of recognition in our experiments– is calculated with Equation (5.8). This overall accuracy provides an indication on the proportion of the asset images that can be reliably segmented.

$$\Omega_a = \frac{\sum_i CM[i, j]}{\sum_i \sum_j CM[i, j]} \tag{5.8}$$

4.2.3. Experimental results and discussion

The developed method is implemented in Matlab on Windows 64bit. The experiment was benchmarked on an Intel(R) Core(TM) i7-3820 CPU @ 3.60 GHz with 64.0 GB RAM and NVIDIA GeForce GTX 400 graphics card. The Matlab pool is used to enable parallel computation by creating jobs on a pool of workers and connecting the pool to the Matlab client. In the following, we present experimental results on each of our datasets.

a. Smart Road results

To fairly report the performance on the Smart Road dataset, not only we evaluate the accuracy by the per-pixel classification rate – which is mainly determined by how well we can label the number of asset categories – but also the average of the per-pixel rates over all the asset categories. Figure 4.14 shows examples of the experimental results. As observed, most parts of these video frames are properly segmented and labeled with their corresponding asset categories.

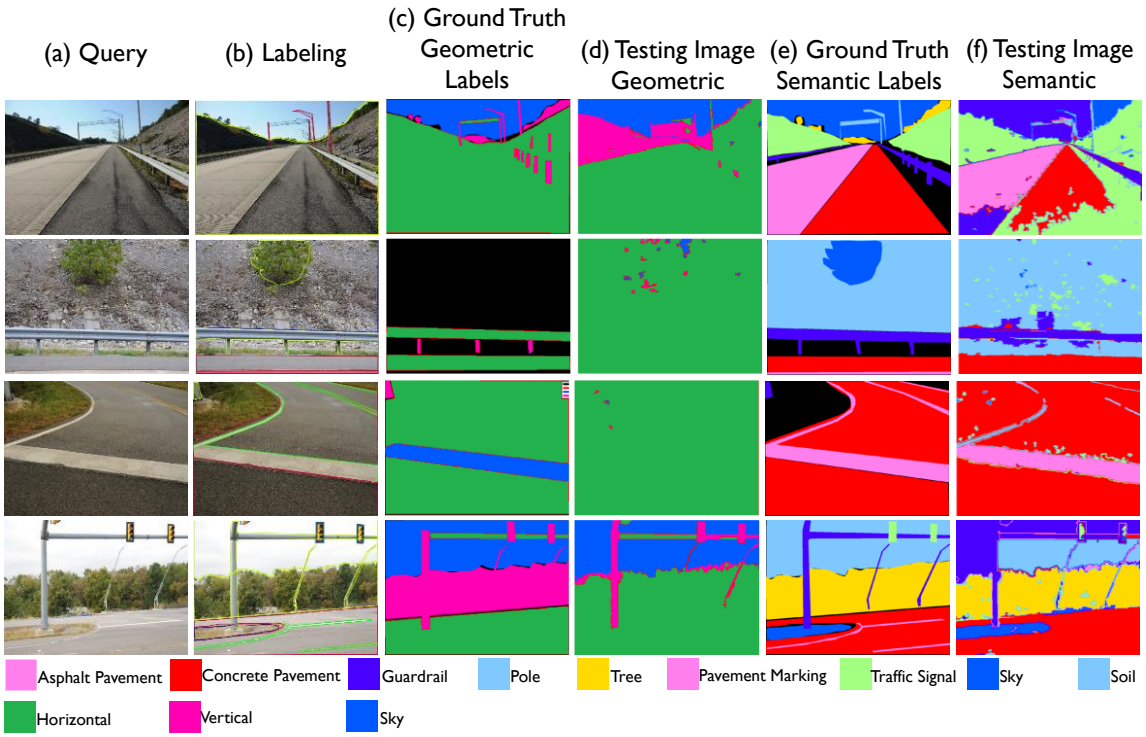

Figure 4.14 Example Results from the Smart Road Testing Dataset

In order to measure the accuracy of recognition on testing images, the outcomes are compared with their corresponding ground truth images. Hence, the color value at each pixel in the segmented imagery is compared with ground truth. The accuracy of segmentation for each category of roadway assets is presented in Table 4.5. The average accuracy is 87.13%.

We also compare our results with those reported in Section 5.1. As shown in Table 4.6, the average accuracy of recognition is increased by 0.38%. This is because mapping geometric labels to semantic labels can increase the accuracy of recognition for asphalt pavement, guardrails, light poles, soil, and pavement marking categories. As reported in Section 5.1 the main confusions for the Semantic Texton Forest method are between asphalt pavement, concrete pavement, and soil categories while adding the geometric information to the semantic context has increased the

accuracy. Our new method also outperforms the computational time for the Semantic Texton Forest method with an order of magnitude, and requires significantly smaller amount of supervision in the training process.

Table 4.5 Accuracy of 2D Video Frame Semantic Segmentation

Asset Labels	Accuracy (Percent)
Asphalt Pavement	89.29
Concrete Pavement	96.69
Guardrail	88.54
Light Poles	78.36
Traffic Signs	96.61
Trees	76.43
Grass	72.69
Soil	90.4
Sky	96.43
Safety Cones	81.38
Traffic Signals	86.03
Pavement Markings	92.76

Table 4.6 Comparison of Segmentation Accuracy for Superparsing and Semantic Texton Forest Method on Smart Road Dataset

Asset Labels	Superparsing	STF	Difference
Asphalt Pavement	89.29	82.58	+6.71
Concrete Pavement	96.69	99.04	-2.35
Guardrail	88.54	85.81	+2.73
Light Poles	78.36	71.77	+6.59
Traffic Signs	96.61	98.05	-1.44
Trees	76.43	78.62	-2.19
Grass	72.69	72.3	+0.39
Soil	90.4	87.3	+3.1
Sky	96.43	98.25	-1.82
Safety Cones	81.38	85.89	-4.51
Traffic Signals	86.03	91.78	-5.75
Pavement Markings	92.76	89.67	+3.09
Average Accuracy	87.13	86.75	+0.38

Figure 4.15 shows the confusion matrix for segmentation of asset categories. The average accuracy of 79% for asset segmentation is achieved which indicate how accurately each superpixel region is segmented in the video frames. Such average accuracy shows 2.5% better on the performance of the new method compared to Semantic Texton Forest based method which reports

76.5% on the same testing dataset. The proposed method shows the best performance on traffic signals, safety cones, and guardrails. As it can be observed, the maximum confusion happens between asphalt pavement and concrete pavement asset categories. This is primarily related to the visual consistency of these two categories. This confusion has been decreased due to adding the geometric information to the semantic context with respect to our previous work using Semantic Texton Forest. Some examples of segmentation results of the same images in both Semantic Texton Forest and superparsing method is shown in Figure 4.16.

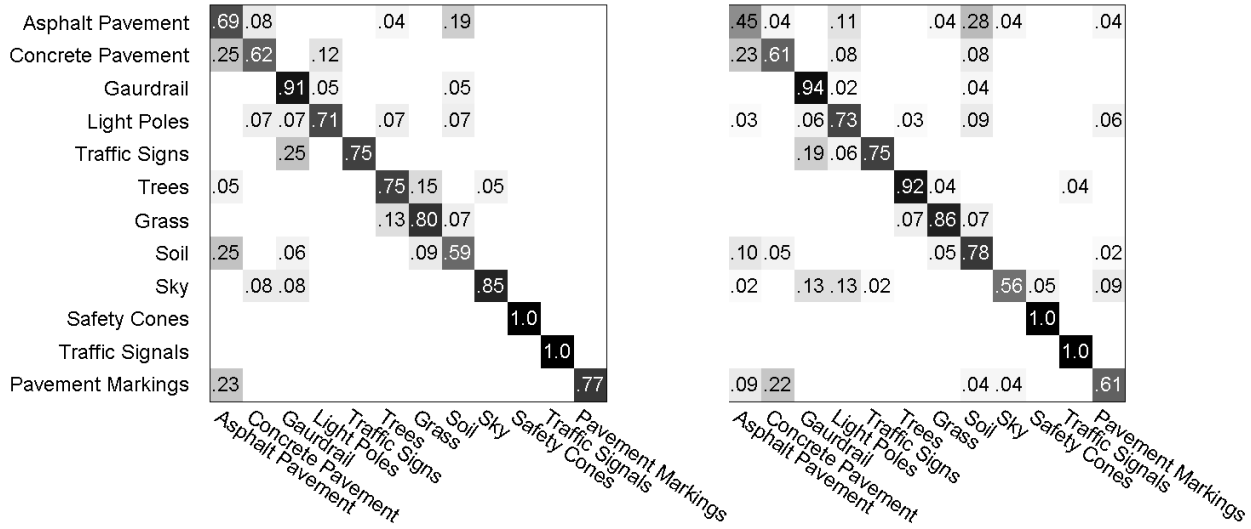

Figure 4.15 Confusion Matrix on the Smart Road Testing Dataset

b. Interstate I-57 results

The second and more comprehensive dataset in our experiment is the I-57 dataset. Figure 4.17 shows several examples of the experimental results on the segmentation of assets. As observed, most parts of these video frames are properly segmented for the expected assets. The final results on the I-57 dataset achieves a classification rate of 88.24%. In order to measure the accuracy of recognition on both training and testing images, we compare the segmented video frames at the pixel level with their corresponding ground truth. Table 4.7 shows accuracy of recognition on training and testing video frames .The accuracy of segmentation is shown through the confusion matrix for segmentation of asset categories. Overall, an average accuracy of 82.02% for region segmentation is achieved.

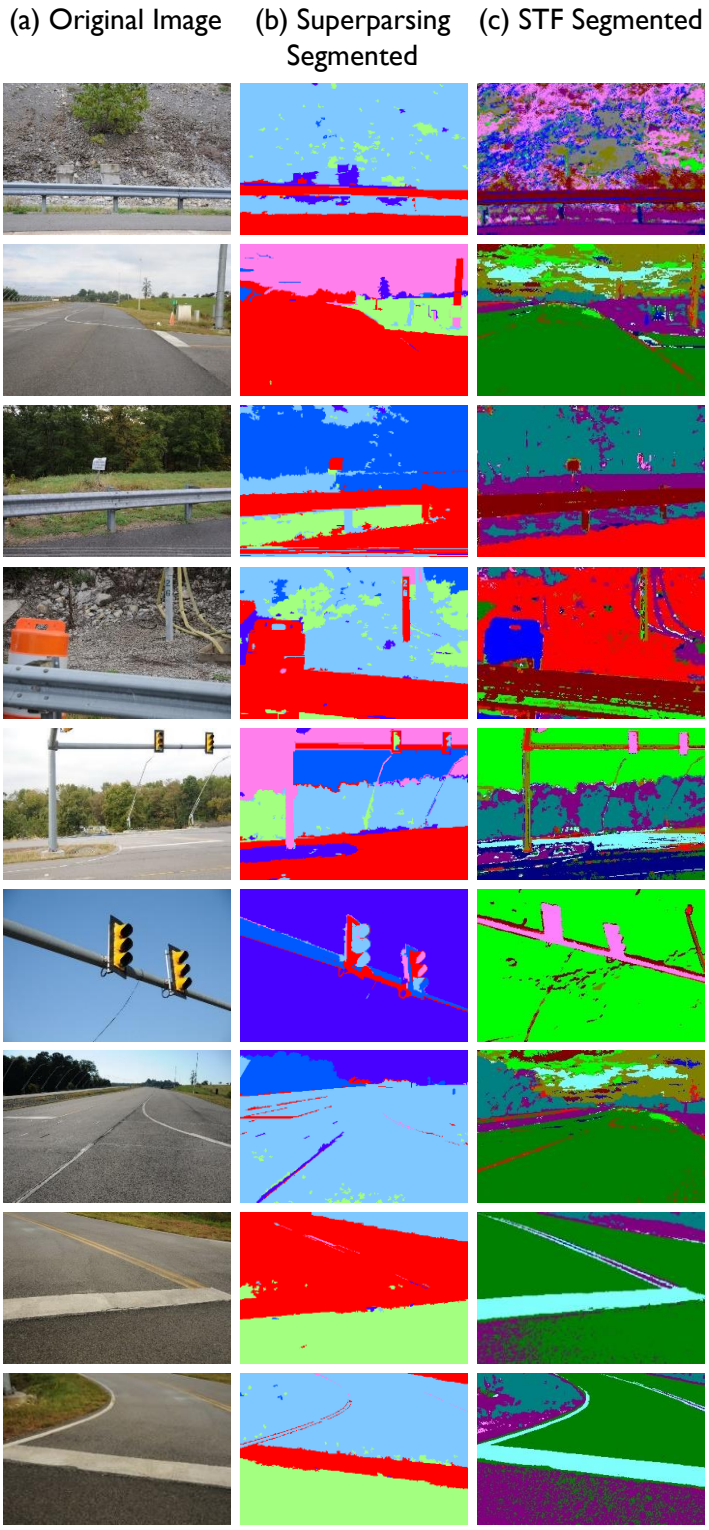

Figure 4.16 Several Examples, Illustrating the Differences Between Superparsing and Semantic Texton Forest Based Methods on the Smart Road Testing Dataset. Although Colors Are Different, All Segmentation Correspond Uniformly Across the Methods

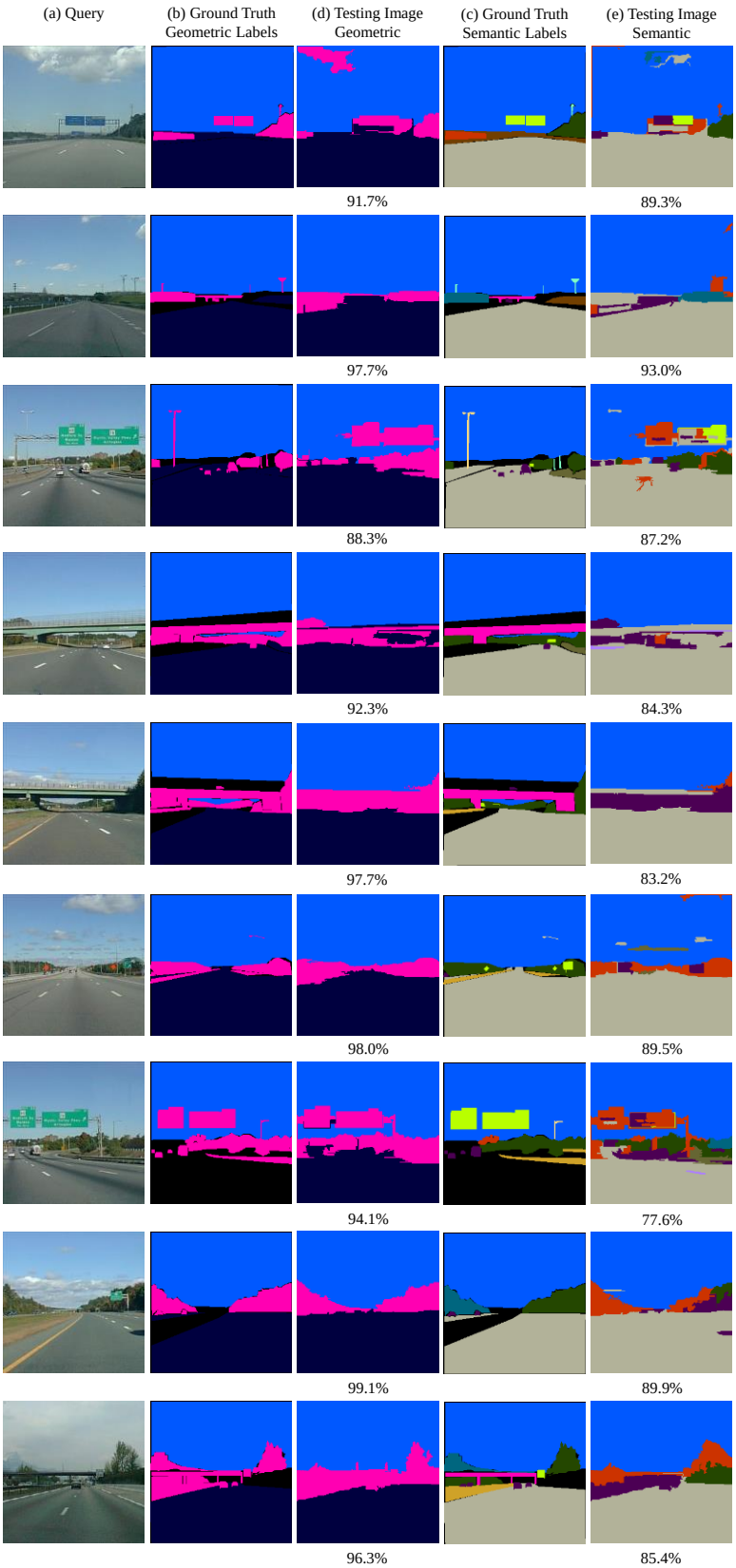

Figure 4.17 Results of Segmentation for Both Geometric Labels and Semantic Labels

Table 4.7 Accuracy of Recognition on I-57 Dataset

Asset Labels	Accuracy (Percent)
Pavement	91.99
Shoulder	93.35
Guardrail	87.17
Pavement Markings	91.22
Light Poles	76.48
Traffic Signs	90.37
Bridges	90.36
Sky	92.52
Others	80.73
Average Accuracy	**88.24**

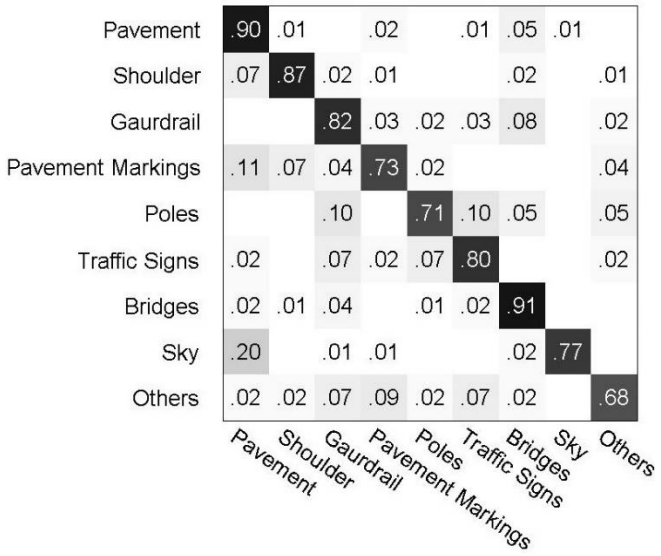

Figure 4.18 Confusion Matrix of Segmentation on I-57 Dataset

c. Video parsing

The video segmentation method was tested on the I-57 dataset which includes frames of video streams. There are a total of 550 labeled frames in the dataset with 347 used for training and 203 for testing. Figure 4.19 shows some examples of video parsing results for continuous frames of video stream.

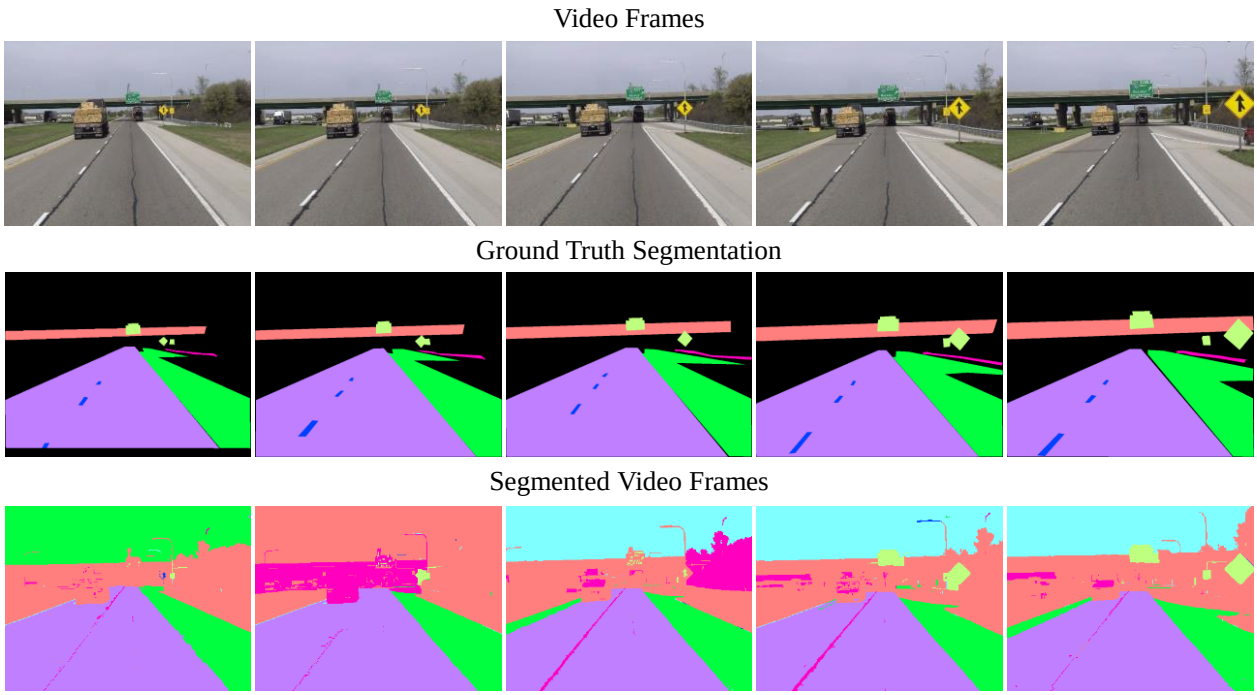

Figure 4.19 Results of Video Parsing on I-57 Dataset

4.3. Evaluation of Multi-Class Traffic Sign Detection and Classification Methods for U.S. Roadway Asset Inventory Management

4.3.1. Data collection and setup

For evaluating the performance of the multi-class traffic sign detection and classification, the experimental dataset was collected along the US-460 and US Interstate 57 by the State of Illinois' Department of Transportation. This dataset is used for both training and testing purposes with split of 70 percent for training and 30 percent for testing. This dataset is released publicly through http://raamac.cee.illinois.edu/aca for other researchers to further develop and validate new algorithms.

The dataset contains different categories of traffic signs based on the signs' message which are annotated manually for training and testing purposes. Training frames are cropped to contain only single traffic sign. In order to create a comprehensive dataset with varying viewpoints, scale, illumination changes, and intra-class variability the videos were collected in different weather condition and on both highway and roadway. To increase negative samples which are needed for training AdaBoost classifier, we also added 16,000 negative samples of typical backgrounds of roadways and highways. The negative images for each binary classification process includes both

positive examples of other categories of traffic signs and also generic roadway and highway backgrounds. Table 4.8 shows the specification of the training and testing datasets.

Table 4.8 Specification of Our Released Traffic Sign Dataset

Type	Color	Dataset	Positive	Negative	Sign Message	
Warning	Yellow	Training Testing	1523 653	6,174 2,658	Warning	
Regulatory	White, Blue, Green	Training Testing	5924 2539	7,353 3,311	Regulatory, Direction (including mile markers)	
Stop Sign	Red	Training Testing	164 71	2,228 3,240	Always means Stop	
Yield	Red	Training Testing	109 48	2,245 3,263	Yield, Slow down, Prepare to Stop	

To achieve the best performance, we conduct experiments with several spatial scales of (0.75, 1.00, 1.25) of the template spatial resolution and consider a 6.67% shift among observed pixels (i.e., window overlap) to find the best candidates for the traffic signs. The results of different sliding window size for different types of traffic sign has been tested. The 64×64 pixel image patches, as shown in Table 4.9, have the minimum FN rates and maximum FP rates among all other sliding window sizes.

Table 4.9 Specification of Our Released Traffic Sign Dataset

	False Negatives Rate			False Positives Rate		
Scale Factor	0.75	1.00	1.25	0.75	1.00	1.25
Warning	0.06%	0.06%	3.64%	13.68%	14.19%	36.76%
Regulatory	3.67%	2.17%	6.48%	11.19%	19.23%	21.95%
Stop Sign	0.00%	0.00%	9.23%	23.08%	30.77%	23.08%
Yield	0.00%	0.00%	4.49%	14.61%	29.21%	38.20%

4.3.2. Performance evaluation metrics

The most straightforward measurement used in the literature is the TP rate. However, even if all the traffic signs are detected, a method is not necessarily perfect. In other words, the ratio of FPs must also be taken into account. If the amount of FPs is too high, the classifier will handle a lot more data than it should, and as a result the overall system speed is degraded. Thus, both TP and FP rates should be accounted simultaneously.

Hence, to quantify and benchmark the performance of different methods, we plot the Precision-Recall graphs. This evaluation metric is extensively used in the Computer Vision community. To facilitate comparing the overall average performance of the variations of proposed approaches over a particular category of traffic signs, individual detection class *precision* values are interpolated to a set of standard *recall* levels (0 to 1 in increments 0.1). Precision is the fraction of retrieved samples that are relevant to the particular classification, while recall is the fraction of relevant samples that are retrieved. Precision and recall are calculated as shown in Equation (5.10):

$$\begin{aligned} \mathrm{Pr}ecision &= \frac{TP}{TP+FP} \\ \mathrm{Re}call &= \frac{TP}{TP+FN} \end{aligned} \tag{5.10}$$

The particular rule used to interpolate precision to recall level *i* is to use the maximum precision obtained from detection class for any recall level greater or equal to *i*. For each recall level, the precision is calculated, then the values are connected and plotted to form a curve.

4.3.3. Experimental Results and Discussion

As a proof of concept, we prototyped these methods in Matlab on a Windows 64bit workstation. The performance of our implementation was benchmarked on an Intel(R) Core(TM) i7-3820 CPU @ 3.60 GHz with 64.0 GB RAM and NVIDIA GeForce GTX 400 graphics card. The specification of the methods used are presented in Table 4.10.

Table 4.10 Specification of Our Methods

Method	Properties
1. Haar	64×64 pixel base templates for the detection windows
2. HOG	Linear gradient [-1;0;1] voting into 8 orientation bins in 0-180˚ L2-normalized blocks with 4 cells containing 8×8 pixels
3. HOG + Color	Linear SVM Classifiers with $C=1$ Linear, Polynomial, and RBF SVM classifiers

The HOG templates which are trained separately for each type of traffic signs are shown in Figure 4.20. In this figure, the second row shows how a computer sees the same training images.

The bottom row shows a standard visualization where shadows are removed, fine details are lost, and the image is noisier, but the template of each sign is preserved.

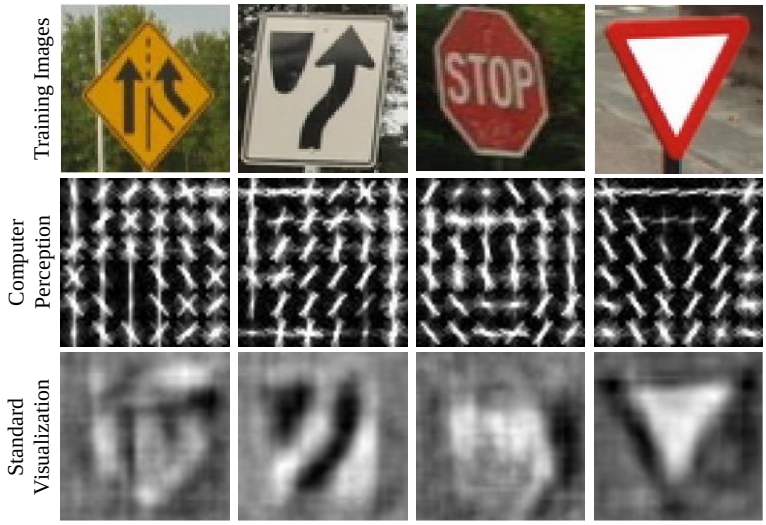

Figure 4.20 Examples of the Visualization for HOG Feature Space: First Row Shows the Training Images; Second Row Shows How a Computer Sees the Same Images. The Bottom Row Shows a Standard Visualization

Figure 4.21 shows examples of the testing images, in addition to visualization of their HOG and color descriptors. Here, the HOG descriptors remain non-sensitive to variations in lighting conditions. The choice of Hue and Saturation color values also guarantee that our representations are invariant to changes of scene brightness.

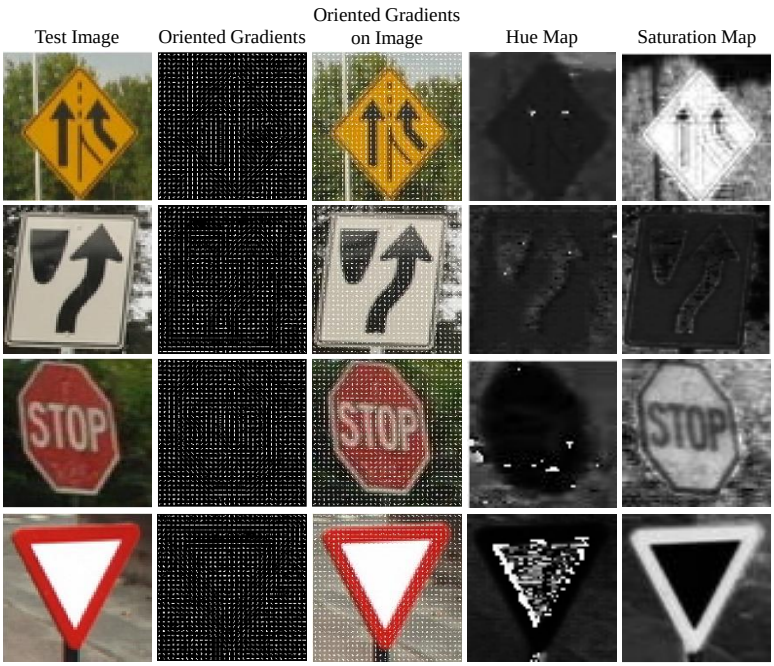

Figure 4.21 Examples of Testing for Different Types of Traffic Signs

In the first phase of validation, experiments were conducted to test the performance of the multiple classifiers while accounting for the impact of the parameters of the sliding window overlap and also the spatial resolutions. Figure 4.22 shows different cases on the performance of the classification. The first row shows TPs which are correctly identified. The second row shows multiple detections due to the impact of the sliding windows on the detections. Third row here shows incorrectly classified traffic signs which are FPs. And the last row shows the incorrectly rejected signs which are FNs.

Figure 4.22 Examples of TP, FP, and FN of Sliding Window with Size of 64×64 Pixels for Detection and Classification of Traffic Signs

The precision and recall metric is used to measure the performance of detection and classification for candidate window size 64×64 pixels for different types of traffic signs. Precision and recall graphs of different detection approaches and different classifiers are shown in Figure 4.23.

As observed, the HOG+C method improves the performance of detecting traffic signs compare to Haar-like feature and HOG. In particular it achieves higher precisions in higher recall values. The average precision, recall, and accuracy of different types of traffic signs and approaches are shown in Table 4.11. The accuracy of classification is calculated based on $\frac{TP+TN}{TP+TN+FP+FN}$.

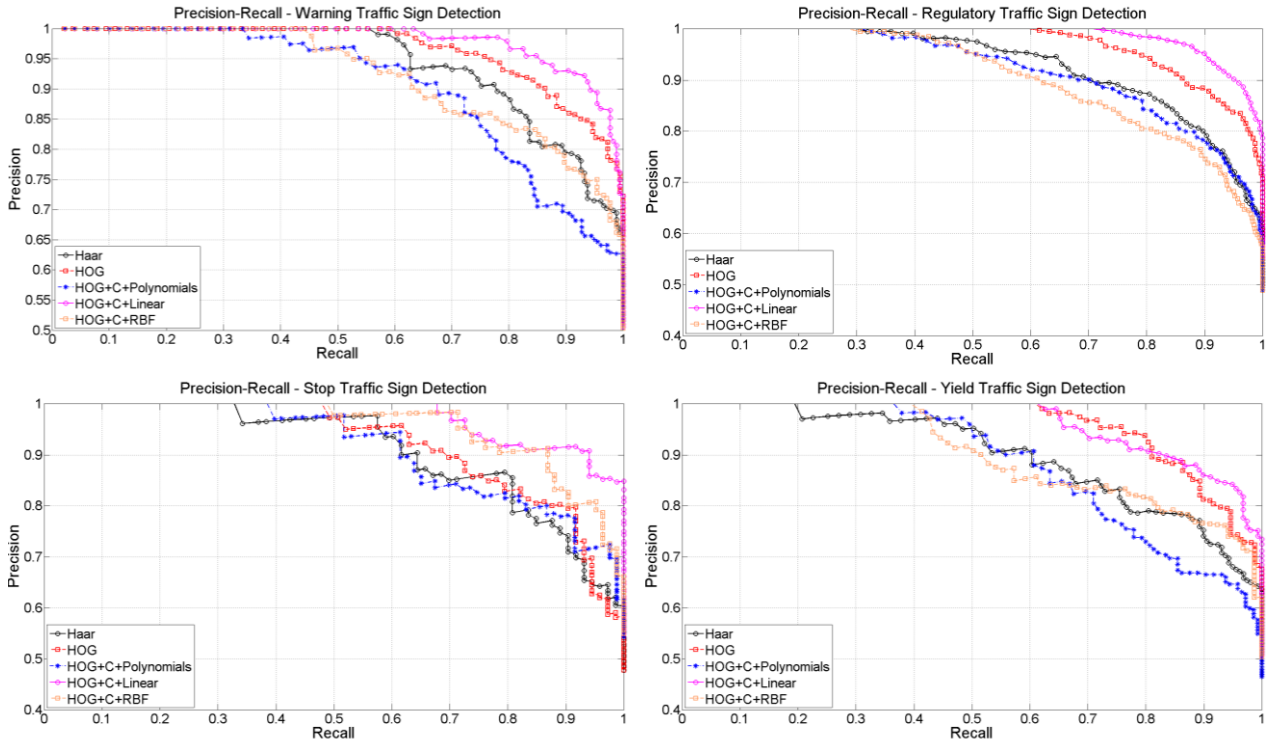

Figure 4.23 Precision-Recall Graphics on the Performance of Our Three Methods for Multi-class US Traffic Sign Detection and Classification

Table 4.11 Precision, Recall, and Accuracy of Different Types of Traffic Signs Considering Different Approaches

Approach	Precision (%)			Recall (%)			Accuracy (%)		
	Haar	**HOG**	**HOG+C**	**Haar**	**HOG**	**HOG+C**	**Haar**	**HOG**	**HOG+C**
Warning	85.81	95.84	**97.19**	99.93	99.87	**99.93**	85.75	95.72	**97.12**
Regulatory	80.77	92.06	**92.85**	97.38	96.14	**98.18**	79.05	88.78	**91.28**
Stop Sign	69.23	73.85	**92.31**	100	100	**98.36**	69.23	73.85	**90.91**
Yield	70.79	98.88	**100**	100	100	**100**	70.79	98.88	**100**
						Average	76.20	89.31	**94.83**

With the computer configuration explained earlier, the new method of HOG+C and multiple one-vs.-all linear SVM also outperforms the other methods in terms of its computational time. The comparison of computational time for both training and testing of each approach for different types of traffic sign are summarized in Table 4.12. It also shows the average for training and testing time per video frame for each method in seconds.

Using HOG+C for the detection and classification of multiple categories of traffic signs has a higher accuracy compared to the Haar and HOG as it leverages both shape and color information in a principled way. As it shown in Table 4.11, due to the distinct colors of traffic

signs, adding the color information and forming HOG+C histograms improves the performance of HOG by 5.5%. In all of the traffic sign categories HOG+C has higher accuracy and in some cases such as Stop sign it improves performance by 23% with respect to Haar and 18.5% with respect to HOG. The reason of this improvement is the special shape and color of stop sign which is unique among all the traffic signs. Figure 4.24 shows several examples from the testing datasets were the performance of detection, 2D localization, and classification at multiple scales and under different roadway and highway background are demonstrated. For more results, readers are encouraged to review the companion video.

Table 4.12 Computational Time of Different Types of Traffic Signs Considering Different Approaches

Type	Dataset	# Images	Haar	HOG	HOG+C
Warning	Training	1,523	3 days	1 day	1.5 days
	Testing	653	1.8 hr	0.5 hr	1 hr
Regulatory	Training	5,924	10 days	3 days	4 days
	Testing	2,539	6.5 hr	2.25 hr	3.5 hr
Stop Sign	Training	164	0.5 days	3 hr	4 hr
	Testing	71	0.25 hr	0.1 hr	0.15 hr
yield	Training	109	0.5 days	3 hr	3.5 hr
	Testing	48	0.15 hr	0.08 hr	0.1 hr
Average per Video Frame	**Training**		**156.68 s**	**49.24 s**	**64.91 s**
	Testing		**9.46 s**	**3.26 s**	**5.18 s**

Figure 4.24 Several Examples of Successful Multi-class Traffic Sign Detection, 2D Localization, and Classification

4.3.4. Discussion on the proposed research and challenges

This study presented the first comprehensive video frame data set for 2D detection of US traffic signs and mile markers and to the best of our knowledge is the first evaluation of dominant shape and color object detection algorithms which is implemented at multiple scales for detection multiple categories of traffic signs. Because real-world video frames are used, our dataset exhibits large variations in color and also illumination conditions (i.e. sun light, cloudy weather, shadow). In the absence of a color normalization and/or illumination compensation of the input images, adding color to HOG seem capable of largely making up for these variances. The average accuracy for detection of warning, regulatory, stop, and yield traffic signs using the HOG+C method are 97.12%, 91.28%, 90.91%, and 100% respectively. As observed in Table 4.12, implementing both HOG and HOG+C on graphic process unit (GPU) or using multi-core CPU can make real-time traffic sign detection and classification.

Even though this work is mainly focused on detection, 2D localization, and classification of different types of traffic signs, it has several other applications:

1) *Collecting comprehensive sign inventory:* Full inventory on the types of existing traffic signs, together with their location information.
2) *Roadway maintenance*: control the presence and condition of traffic signs along all roadways as opposed to only major roadways and highways.
3) *Driver assistance systems*: assist the drivers by informing them on current road restrictions, speed limits, and warnings.
4) *Intelligent autonomous vehicles*: Provide situational awareness to navigation of intelligent autonomous vehicles.

4.4. Mapping Traffic Signs Using Google Street View Images for Roadway Inventory Management

4.4.1. Data collection and setup

For evaluating the performance of proposed method, the multi-class traffic sign detection model was trained using images collected from a highway and many secondary roadways in the

U.S. This dataset –shown in Table 4.13– contains different categories of traffic signs, which exhibiting various viewpoints, scales, illumination, and intra-class variability. The dataset and the manually annotated ground truth are used for fine tuning the candidate extraction method and also training the SVM classifiers. The models were trained to classify U.S. traffic signs into four categories of warning, regulatory, stop, and yield signs.

Table 4.13 Specification of the Released Traffic Sign Dataset Used for Training SVM Classifiers

Type	Color	Shape	# of images	
Warning	Yellow	Diamond	2,176	
Regulatory	White, Blue, Green	Rectangle	8,463	
Stop Sign	Red	Hexagonal	235	
Yield	Red	Triangle	157	
Generic Backgrounds			10,000	

In this paper, the data collected from Google Street View API is used purely as the testing dataset. This dataset is collected on 6.2 miles in two segments of U. S. I-57 and I-74 interstate highways (see Figure 4.25).

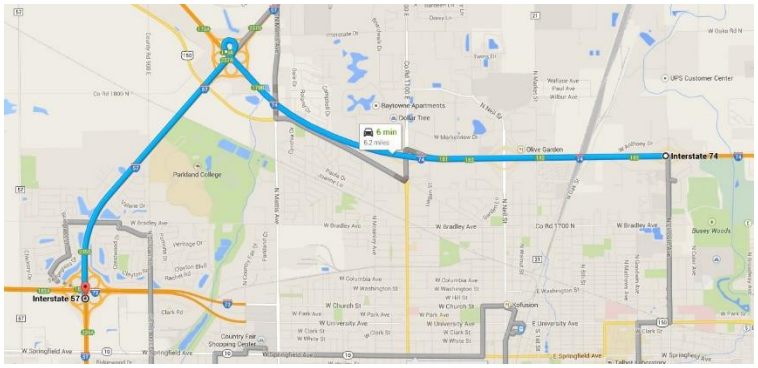

Figure 4.25 Testing Route on I-74 and I-57- 6.2 Miles Long

Google Street View images can be downloaded in any size up to 2048×2048 pixels. Figure 4.26 shows a snapshot of the API with the information and associated URL for downloading the shown image.

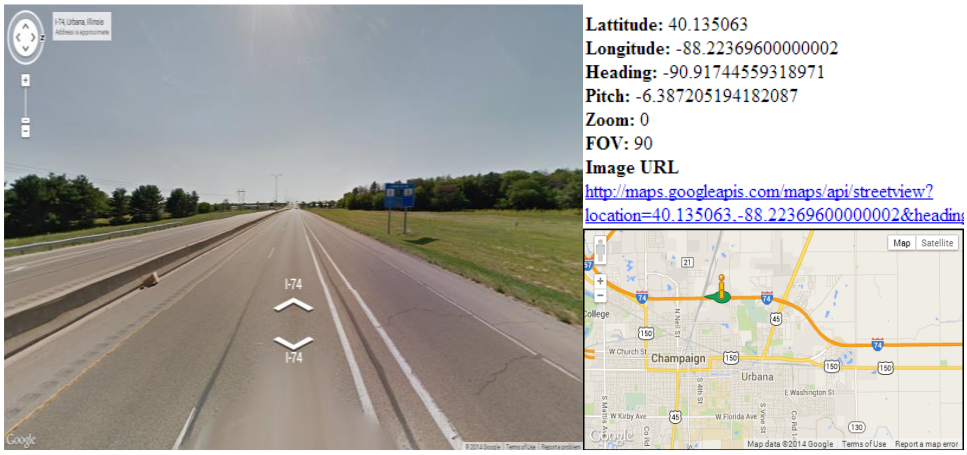

Figure 4.26 Google Street View API

Table 4.14 shows the properties of the HOG+C descriptors. Because of the large size of the training datasets, linear kernels are chosen for classification in the multiple one-vs.-all SVM classifiers. The base spatial resolution of the sliding windows was set to 64×64 pixel with 67% spatial overlap for localization which builds on a non-maxima suppression procedure.

Table 4.14 Parameters of HOG + Color Detectors

HOG Parameters	HOG Values	Color Parameters	Color Values
Linear Gradient	[-1; 0; +1]	Color Channel	Hue and Saturation
Voting Orientation	8 orientations in 0-180°	Number of Bins	6 for each
Normalization Method	L2 Normalization blocks	Normalization Method	L2 Normalization blocks
Number of cells	4	Number of cells	4
Number of Pixels	8×8	Number of Pixels	8×8
Classifier	Linear SVM Classifiers with $C = 1$		

The performance of our implementation was benchmarked on an Intel(R) Core(TM) i7-3820 CPU @ 3.60 GHz with 64.0 GB RAM and NVIDIA GeForce GTX 400 graphics card. The developed system is available online at: http://signvisu.azurewebsites.net/ and the companion video of this manuscript illustrates the functionalities.

4.4.2. Results and Discussion

For validation of proposed detection, classification, and visualization of traffic signs on the Google APIs, ground truth was generated manually. In the first phase of validation, experiments were conducted to detect and classify traffic signs from Google Street View images. Figure 4.27

shows several example results from the application of the multi-category classifiers. As observed, different types of traffic sign with different scales, orientation/pose, and under different background conditions are detected and classified correctly.

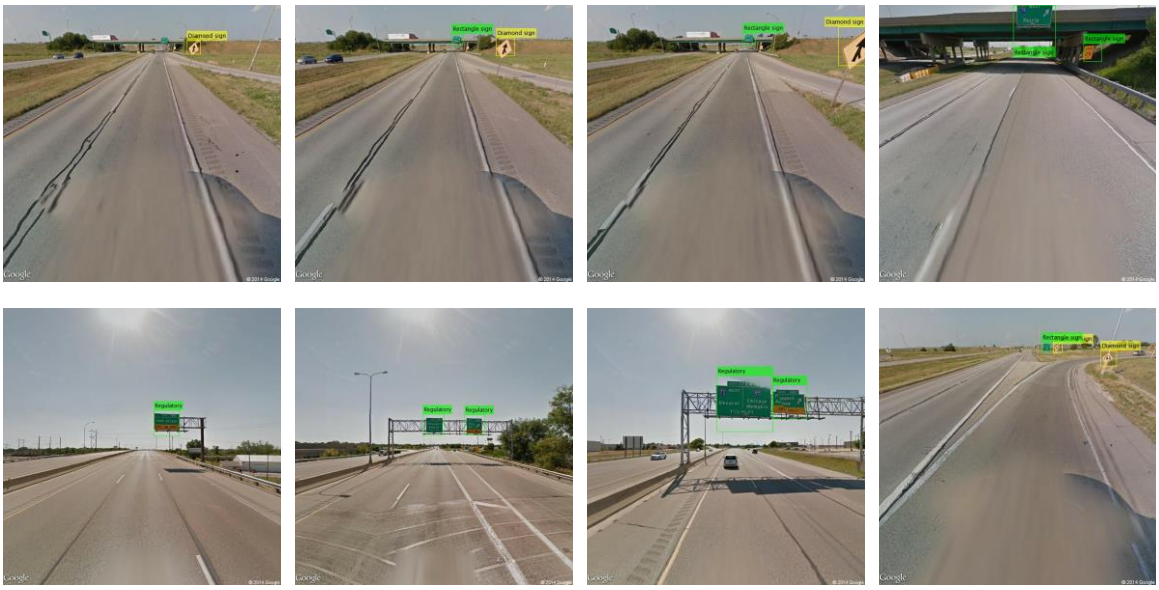

Figure 4.27 Multi-class Traffic Sign Detection and Classification in Google Street View Images

Based on detected traffic signs, a comprehensive database of detected signs is created in which each sign is associated with its most probable location (the image with maximum bounding box size is kept). Figure 4.28 shows an example of data cards which are created for detected signs.

ImageName: 22-pic.jpg
Lat: 40.13504319
Long: -88.22567371
Roadway Number: Interstate 74
City: Urbana
State: IL
ZIP: 61802
Country: USA
Data Collection Date: Jun-12
Category: Traffic Sign
AssetType1: Regulatory
AssetType2:
AssetType3:
X1: 454
Y1: 299
W1: 58
H1: 58
X2: 0
Y2: 0
W2: 0
H2: 0
X3: 0
Y3: 0
W3: 0
H3: 0
Condition of Asset Type 1:
Condition of Asset Type 2:
Condition of Asset Type 3:
URL: http://maps.googleapis.com/maps/api/streetview?pano=mIczTgY_nafio3PAhyZD8Q&size=1024x1024&key=AIzaSyCPhxOoJYHl
latlong: 40.1350431888751 -88.2256737127293
signcount: 1
HeatWeight: 58
latlongifsig: 40.1350457573242 -88.2253706357443
col33: 454
col34: 299
col35: 58
col36: 58

ImageName: 23-pic.jpg
Lat: 40.13504190
Long: -88.22582525
Roadway Number: Interstate 74
City: Urbana
State: IL
ZIP: 61802
Country: USA
Data Collection Date: Jun-12
Category: Traffic Sign
AssetType1: Regulatory
AssetType2:
AssetType3:
X1: 501
Y1: 296
W1: 70
H1: 70
X2: 0
Y2: 0
W2: 0
H2: 0
X3: 0
Y3: 0
W3: 0
H3: 0
Condition of Asset Type 1:
Condition of Asset Type 2:
Condition of Asset Type 3:
URL: http://maps.googleapis.com/maps/api/streetview?pano=eEolcm-iwNMTmbcNjxrCTQ&size=1024x1024&key=AIzaSyCPhxOoJYHlINxYww_3Gn
latlong: 40.1350419046506 -88.2258252512218
signcount: 1
HeatWeight: 70
latlongifsig: 40.1350444730996 -88.2255221742368
col33: 501
col34: 296
col35: 70
col36: 70

Figure 4.28 Data Card for Detected Signs Used for Comprehensive Database of Traffic Signs

Figure 4.29 shows the results from localizing the detected traffic signs: (a) the number of detected signs with the clickable clusters on the Google Map in a section of the I-74, (b) the location markers for the detected signs on Google Earth, (c) the detected sign and its type in the associated Google Street View imagery, and (d) the Google Street View image of the desired location and roadway in which the detected sign is marked. An example of the dynamic heat map for visualizing the most probable 3D location of the detected signs on the Google Earth is shown in Figure 4.29(e). Figure 4.29(f) further illustrate the mapping of all detected signs in multiple locations. The report card for each sign which contain latitude/longitude, roadway number, type of traffic sign, and detection/classification score are shown in this map. These cards facilitate the review of specific sign information in a given location without searching through the large databases. Such spatio-temporal representations can provide DOTs with information on how different types of traffic signs degrade over time and further provides useful condition information necessary for predicting sign replacement plan.

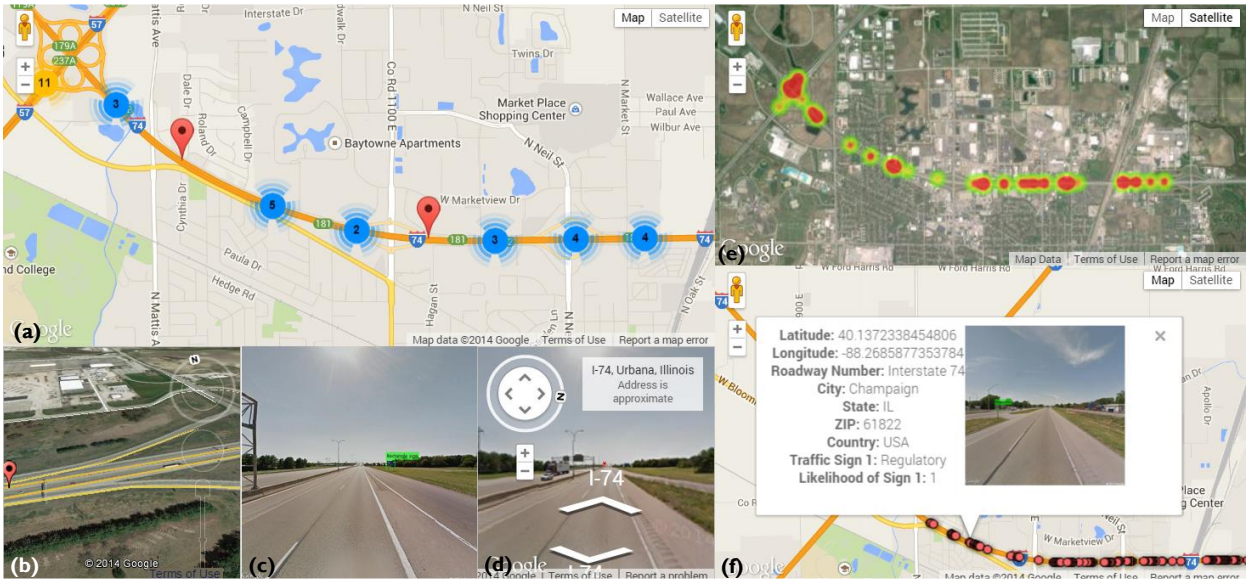

Figure 4.29 Web-based Interface of Developed System; (a) Clustered Detected Signs, Clickable Map; (b) Google Earth View of Sign Location; (c) Detected Sign in Google Street View Image; (d) Street View of Sign Location; (e) Likelihood of Existing Signs on Heat Map; (f) Information on All Detected Signs

To quantify the performance of the detection and classification method, *precision-recall* and *miss rate* metrics are used. Here, precision is the fraction of retrieved instances that are relevant to the particular classification, while recall is the fraction of relevant instances that retrieved:

$$Precision = \frac{TP}{TP + FP} \tag{5.11}$$

$$Recall = \frac{TP}{TP + FN} \tag{5.12}$$

In the precision-recall graph, the particular rule used to interpolate precision at recall level i is to use the maximum precision obtained from the detection class for any recall level greater than or equal to i. Miss rate, as shown in Equation (5.13), shows rate of FNs for each category of traffic signs while FPPW measure the rate of false positives per window of detection. Based on this metric, a better performance of the detector should achieve minimum miss rate. The average accuracy in traffic sign detection and classification using Google Street View images is also calculated using Equation (5.15):

$$miss\ rare = 1 - \ Recall = \frac{FN}{TP + FN} \tag{5.13}$$

$$FPPW = \frac{FP}{FP + TN} \tag{5.14}$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{5.15}$$

The precision-recall, miss rate, and accuracy of detection and localization applied to the I-74 and the I-57 corridors for different types of traffic signs per image and per asset are shown in Table 4.15. Figure 4.30, left to right, shows the Precision-Recall graphs for different types of traffic signs per traffic sign (if it is at least detected from three images) and per image respectively.

Table 4.15 Miss Rate and Accuracy per Image for Different Types of Traffic Sign (Total of 216 Signs)

Accuracy Per	Per Image		Per Asset	
Image	Warning Sign	Regulatory Sign	Warning Sign	Regulatory Sign
Precision	100 %	95.73 %	100 %	87.04 %
Recall	100 %	98.74 %	100 %	95.92 %
Accuracy	100 %	94.58 %	100 %	83.93 %
Miss Rate	0.00 %	1.26 %	0.00 %	4.08 %
FPPW	-	100%	-	100%

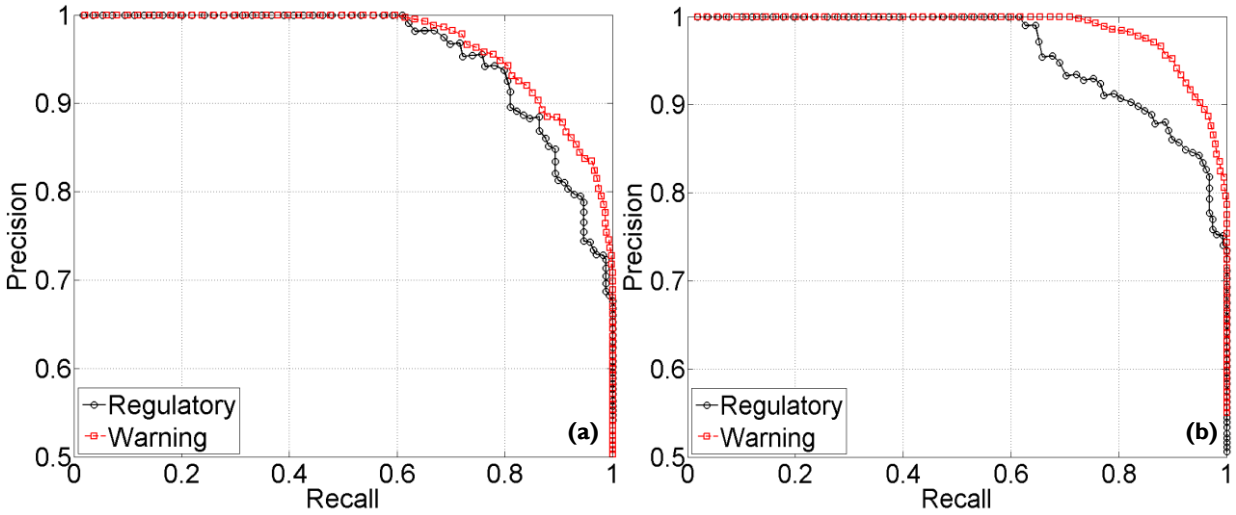

Figure 4.30 Precision-Recall Graph (a) per Asset and (b) per Image for Different Types of Signs

The average miss rate and accuracy in classification among all images is 0.63% and 97.29% and among all types of traffic signs is 2.04% and 91.96%. In other words, only 2.04% signs are not detected in the developed system. Figure 4.31 shows the rate of TPs based on the size of traffic signs in Google Street View Images. As shown, the majority of the traffic signs have been detected using bounding boxes of 40×40 pixels which further validates the choices made in the size of bounding boxes in our developed system.

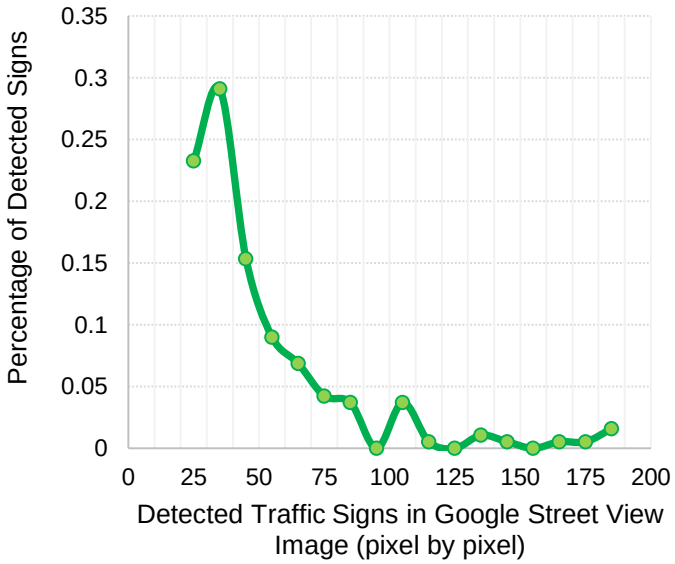

Figure 4.31 Rate of TPs vs Size of Traffic Signs in Google Street View Images

While this work focused on achieving high accuracy in detection and localization of traffic signs, yet the computational time was also benchmarked. Based on the experiments conducted, the computation time for detecting and classifying traffic signs is almost near real-time (5-30 seconds per image). The developed Google API also retrieves and downloads approximately 23 Google Street View images per second. Future study will focus on leveraging Graphic Processing Unit (GPU) to improve the computational time (expected to a high of 10–fold). Even under current computational time, the system allows the Traffic Signs and Marking Division of DOTs to create new traffic sign database while updating existing sign asset locations, attributes, and work orders. The method can also automate the data collection process for ESRI ArcView GIS databases.

4.5. Image-based Retro-Reflectivity Measurement of Traffic Signs in Day Time

For evaluating the performance of the image-based retro-reflectivity measurement of traffic signs in a daytime, several experiments were conducted on four traffic signs including retro-reflective speed limit sign, stop sign, and warning sign, and one non-retro-reflective stop sign. The camera used for data collection is Nikon D300 along with Flash Nikon SB-900.

Figure 4.32 Camera Setup for Data Collection at Different Times of Day and at Different Distances

4.5.1. Data Collection and Setup

HDR images are used for image-based lighting purposes. To derive the omni-directional lighting information, we capture an HDR photograph of a spherical mirror (See Figure 3.28) at six different exposure time (1/15, 1/40, 1/100, 1/250, 1/500, and 1/1000) and plot the response curves using the formulation in Equation (4.22). Figure 4.33 shows these response curves for different

color channels. As mentioned, this calibration is a one-time process for each camera and does not need to be repeated for field experiments.

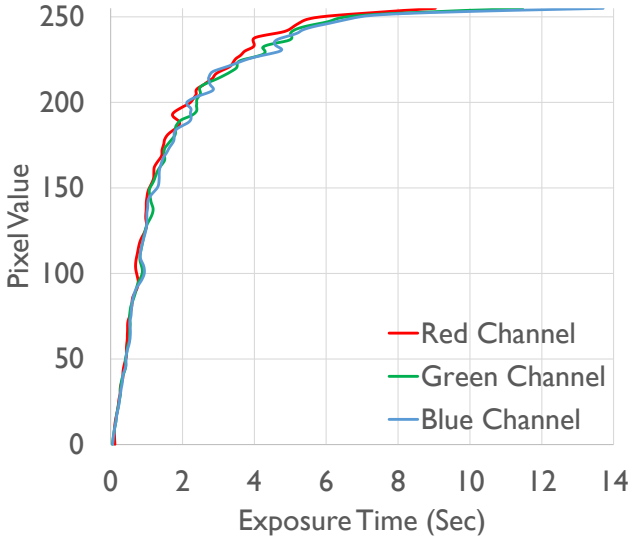

Figure 4.33 Response Curve for Different Color Channels

The experiments were conducted in both sunny and cloudy weather conditions for four times of day: 9:00AM, 12:00PM, 3:00PM, and 6:00PM and at six distances of 25*ft*, 50*ft*, 75*ft*, 100*ft*, 200*ft*, and 250*ft*. Figure 4.34 shows these images. In these experiments, the camera was facing East, which represents the most extreme measurement condition. As a result at 9:00AM, the sun was in front of the camera and at 6:00PM the direction of sun light and flash were the same. Camera setting for data collection is shown in Table 4.16.

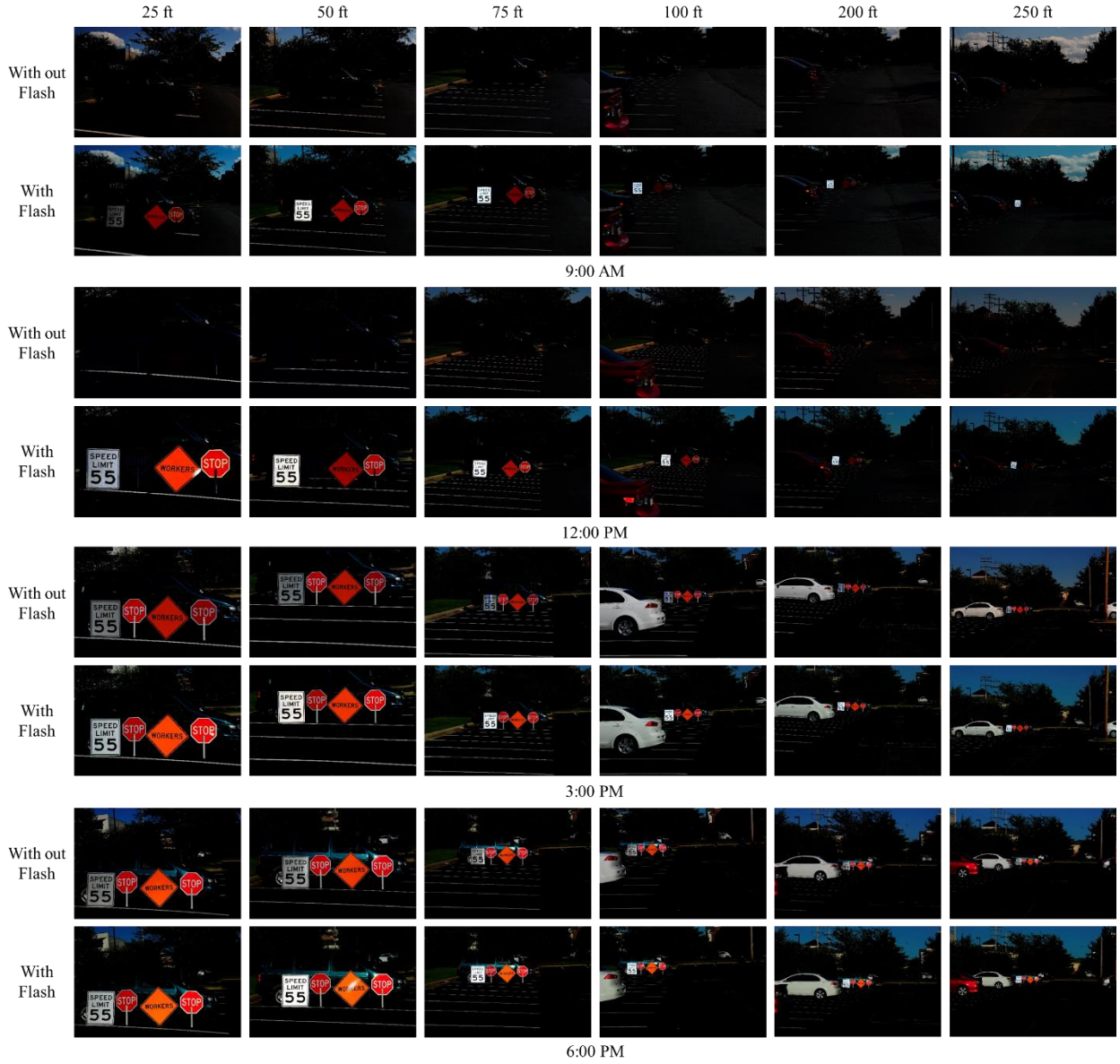

Figure 4.34 Images Collected at Different Times of Day and Different Distances

Table 4.16 Camera Setting for Data Collection

Flash Power	ISO	Exposure Time	Shutter Speed	Focal Length
1/1	L1	f/16	400 ms	70mm

4.5.2. Performance Evaluation

In order to compare the performance of our technique, we benchmarked the as-is retro-reflectivity of our traffic signs at Illinois Department of Transportation's Bureau of Materials and Physical Research in Springfield, IL. At this facility, the retro-reflectivity of all traffic signs is

measured according to ASTM E810-03 guidelines. More specifically, a three-axes goniometer is used to measure the coefficient of retro-reflection on retro-reflective sheeting based the coplanar geometry. The setup, as shown in Figure 4.35, involves the use of a light projector source, a receiver, a device to position the receiver with respect to the source, and a test specimen holder in suitable darkened area. The specimen holder is separated from the light source by 50*ft*. The general procedure involved is to determine the ratio of the light retro-reflected from the test surface to that incident on the test surface. The results of this measurement are shown in Table 4.17 where blank represents measurement when there's no traffic sign (could be interpreted as the measurement tolerance for the 3-axis goniometer).

Figure 4.35 Measuring the Retro-reflectivity Using 3-Axis Goniometer Based on ASTM E810-03

Table 4.17 Results of Ground Truth

Traffic Sign	Speed Limit Sign	Warning Sign	Retro-reflective Stop Sign	Non-retro-reflective Stop Sign	Blank
Retro-reflectivity *(cd/lx*m^2)*	147.625547	13.21307	17.486476	0.115547	0.0503

4.5.3. Results

To check the reliability of our method, several experiments have been carried out. Four different traffic signs with different levels of retro-reflectivity were used to test the performance of our image-based method for retro-reflectivity measurement in daytime. Figure 4.36, Figure 4.37, and Figure 4.38 show the results of image-based retro-reflectivity measurement for speed limit, warning, and retro-reflective stop signs respectively. The retro-reflectivity of each sign at different times of day and for different distances are measured in *cd/lx*m^2* and are compared with the ground truth (measured based on ASTM E810-034). The measurement values are shown under each image in these Figures. In most cases, the measurement shown are above the ground truth values shown in Table 4.17. In a few cases shown with asterisk, the measured retro-reflectivity

numbers are below the ground truth. These are mainly due to distance and the timing of the day used for data collection.

To obtain the criteria for best performance, we examined the impact of distance and lighting condition (the timing of the experiment) on the accuracy of measurement, as follows:

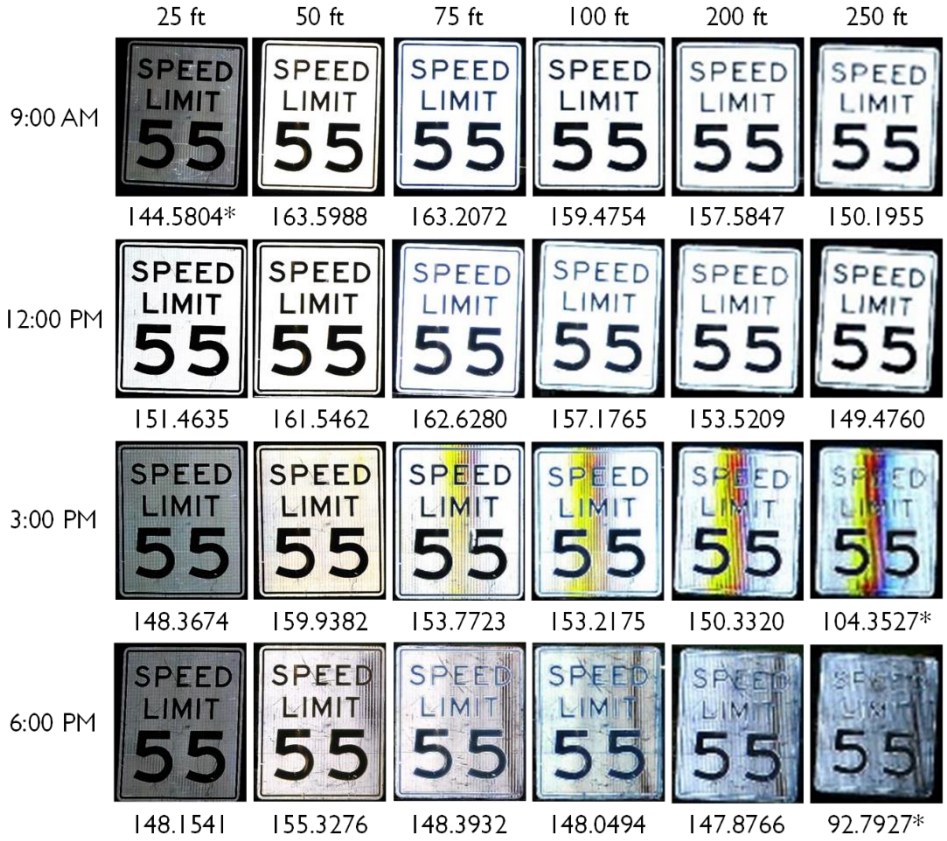

Figure 4.36 Results of Image-based Retro-reflectivity Measurement for Speed Limit Sign

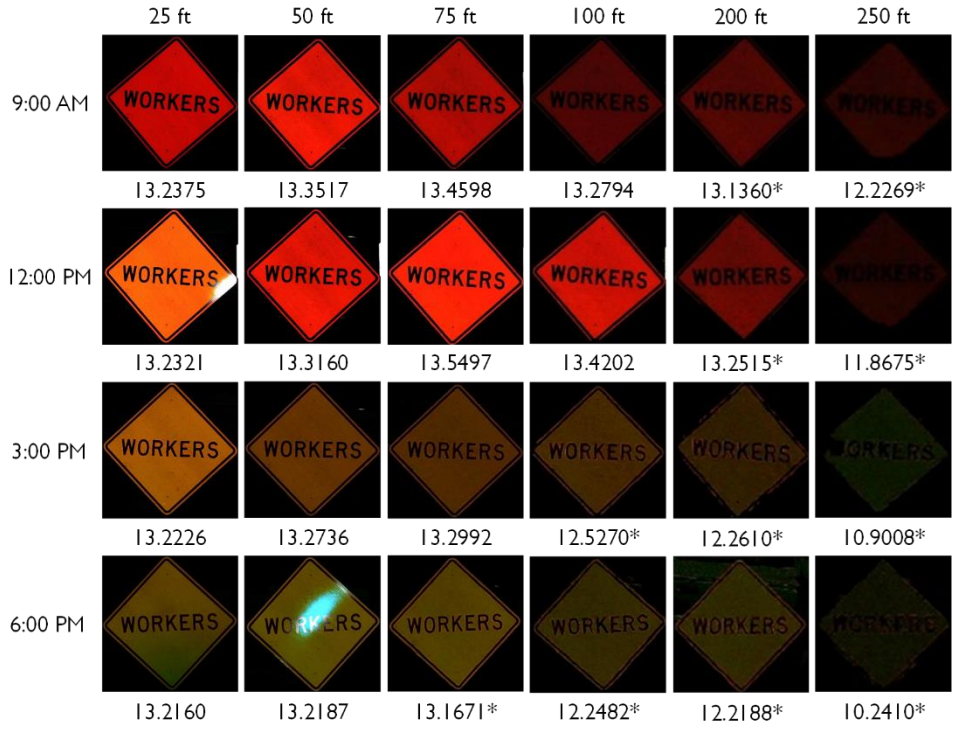

Figure 4.37 Results of Image-based Retro-reflectivity Measurement for Warning Sign

Figure 4.38 Results of Image-based Retro-reflectivity Measurement for Retro-reflective Stop Sign

a. Impact of Time

Figure 4.39 compares the retro-reflectivity measurements for different traffic signs at different times of day. Compared to ground truth, our method with camera facing East (worst measurement condition) works properly at 9:00AM, 12:00PM, and 3:00PM and for all distances less than 75*ft*. However, the performance of our method is decreased as the measurement time approaches the timing of the sunset. The decrease is due to the alignment of the sunlight direction and the camera flash light source direction (West-East in our setup). As shown in most extreme case (6:00PM), when these directions are aligned, our method with current hardware setting is not resulting in accurate measurements.

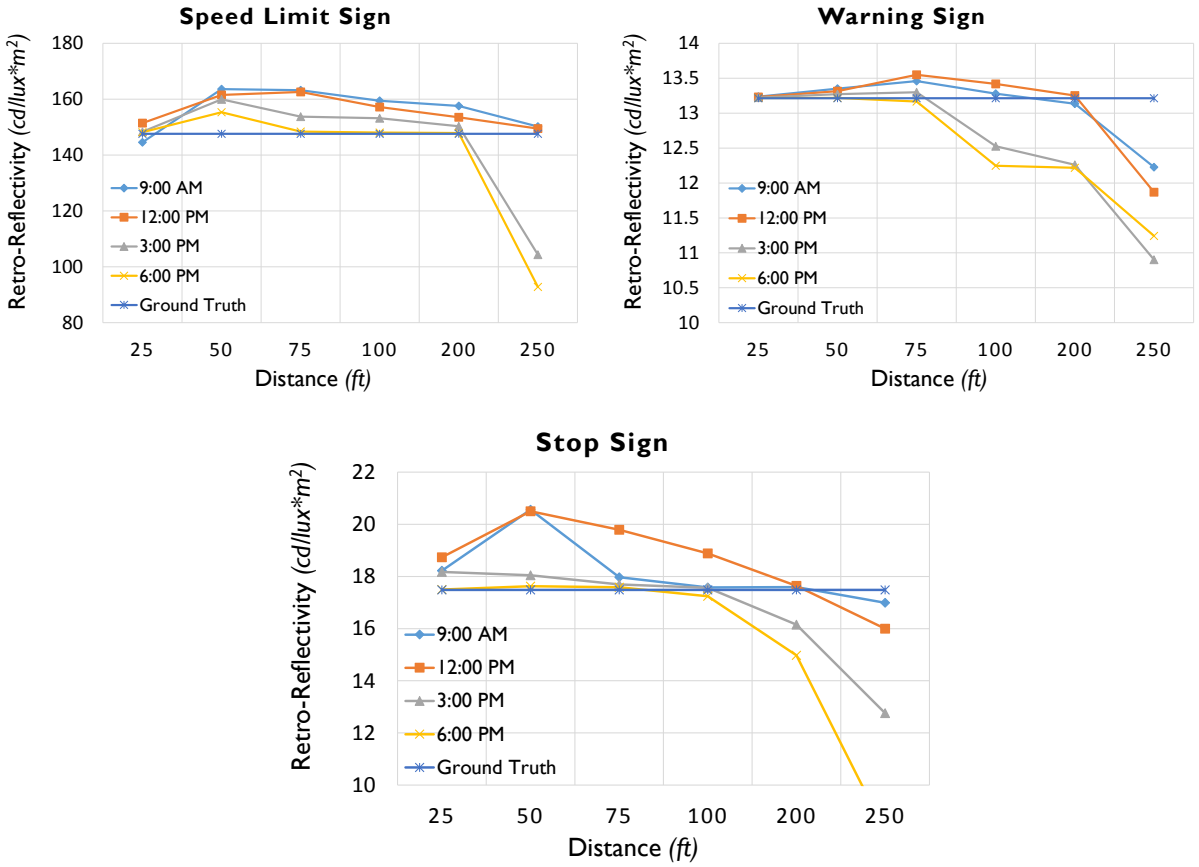

Figure 4.39 Impact of Time on Image-based Retro-reflectivity Measurement for Different Types of Traffic Signs at Different Distances

b. Impact of Distance

Figure 4.40 compares the measurement of retro-reflectivity for different traffic signs at different distances. As shown, our method works pretty well for distances less than 75*ft* in all the times during the day. However, our method is not robust enough for distances above 100*ft* and

more especially in more extreme light conditions (3:00PM and 6:00PM). Nevertheless, our method with current hardware setting is capable of measuring the retro-reflectivity of traffic signs for distances less than 75*ft* and at all times in the day. The accuracy and granularity of our measurements show that our technique complies with FHWA requirements.

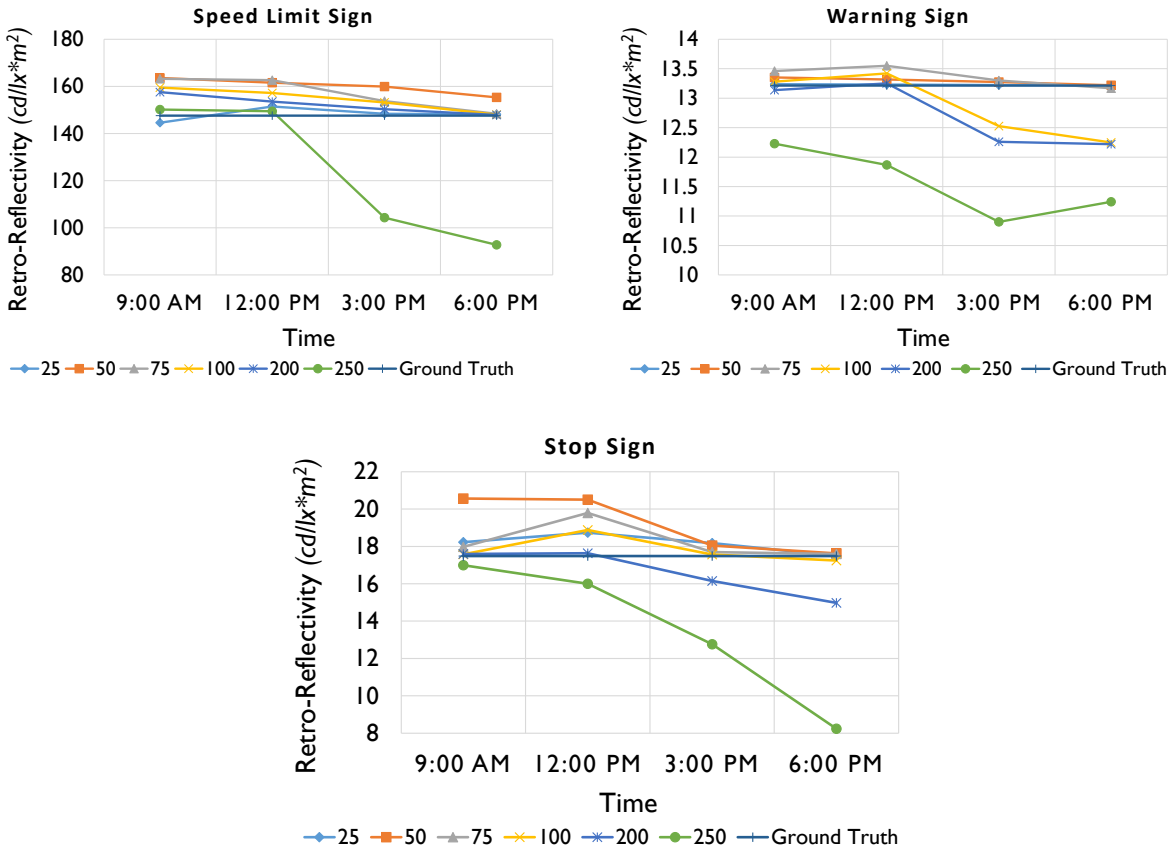

Figure 4.40 Impact of Distance on Image-based Retro-reflectivity Measurement for Different Types of Traffic Signs at Different Times

4.5.4. Discussion

By contrasting the measurements obtained from our method with lab measurements (ground truth), our method shows an accuracy of 88.8%+ in terms of correctly clarifying the measured retro-reflectivity levels as accepted or not. Table 4.18 summarizes these statistics for different types of the signs. For speed limit sign with ground truth retro-reflectivity of 147.625547 *(cd/lx*m^2)*, our method does not produce accurate results at the distance of 250*ft* and at any time between 3:00 to 6:00PM. The proposed method also does not result in accurate measurements on warning sign, at 6:00PM for distances more than 75*ft*, and at time of 3:00PM for distances more than 100*ft*. The same situation applies to retro-reflective stop sign. Based on current hardware and

software settings, our proposed method can measure retro reflectivity of signs at 97.70% accuracy for all distances between 25*ft* and 100*ft* and at any time between 9:00AM to 3:00PM.

Table 4.18 Performance of Proposed Method

Retro-Reflectivity Measurements	**For all distances and at all times**			**For all distances between 25*ft* and 100*ft* and any time between 9:00AM and 3:00PM**		
	Speed Limit Sign	**Warning Sign**	**Stop Sign**	**Speed Limit Sign**	**Warning Sign**	**Stop Sign**
Ground Truth *(cd/lx*m^2)*	147.6255	13.2131	17.4865	147.6255	13.2131	17.4865
Mean	149.3761	12.8384	17.2511	154.4429	13.2641	18.6437
Standard Deviation	16.344	0.71356	2.46884	9.3727	0.4896	1.7363
Accuracy	90.11 %	91.76 %	84.54%	99.72 %	96.68 %	96.69 %

Compared to the current practice of using the retro-reflectometer where only a few measurements are conducted (typically four point-level measurements on sign background, and four point-level measurements on sign text), our method considers the entirety of the traffic sign surface and results in a more comprehensive retro-reflectivity measurement. This is important as traffic signs exhibit heterogeneous deterioration rates, and point-level measurements may not be the best representatives for the entirety of the sign surface. Considering current practical limitations – even when the most accurate retro-reflectometers are used– our method at 97.70% accuracy shows significant promise for large scale applications.

CHAPTER 5. SUMMARY, CONCLUSION, AND PATH FORWARD

Data integration is very important as agencies move toward more global asset management approaches to comprehensively manage different types of transportation assets. Nonetheless, current systems in use only focus on detection of one type of asset, and thus cannot be employed to perform condition assessment for many different types of assets. The predominant contemporary approach to object detection and recognition is to search for representative features of an object rather than searching for the object directly. Development of object/feature recognition algorithms is strongly recommended. Due to minor differences between objects of the same type known as intra-class variation, a new template would have to be trained for every potential object. Being familiar with applicability of discussed technologies can provide asset management and condition assessment researchers and evaluators with an accurate and comprehensive database of all types of assets, allowing the former to build on this research towards the automation of condition assessment, and the latter to make informed decisions based on condition assessment and the best timing and strategies for maintenance. By utilizing the proposed technologies, highway agencies would not only obtain low-cost, accurate, and frequent condition data. But this consistent data also can be used to set the discrete representation of conditions of the low capital assets and formulate the deterioration rates for these assets. Consequently, that would allow better investment planning for low-capital assets. It is recommended that road administration consider the development and implementation of an integrated 2D and 3D vision based system. Such approach can provide better data analysis and decision making to achieve more efficient and higher level of services. Vision-based approaches through inexpensive, sustainable, and easy to install solutions can support automated detection, localization, and visualization of various types of assets. Timed comparison between human-based and vision-based technology for collecting and assessing the condition of high-quantity low-cost roadway assets is directly dependent on the processing and analysis methods employed. The time to analyze the data using the technology is faster than traditional manual collection; however, a significance amount of time is used to process the data to a visible viewing format. Moreover, most existing approaches may not work well in weather impact days, different illumination and visibility, and damaged, misaligned or rotated assets. A critical part of the evaluation process is the availability of standard datasets which different vision based algorithms can be evaluated.

5.1. Summary

This dissertation presented several parts of my research toward recognition and 3D reconstruction of high-quantity low-cost roadway assets specifically traffic signs for enhanced condition assessment.

5.1.1. Segmentation and Recognition of Roadway Assets using Image-based 3D Point Clouds and Semantic Texton Forests

In this dissertation, we presented a new automated and integrated image-based roadway asset 3D reconstruction and 2D segmentation method. Experimental results with an average accuracy of 76.50% and 86.75% in segmentation and pixel-level recognition of 12 types of asset categories reflect the promise of the applicability of this approach for segmentation and recognition of roadway assets from image-based 3D point clouds. The low-cost and accuracy of this method along with the high safety associated with its application can minimize several challenges associated with current manual and subjective data analysis and/or the computer vision systems that are currently in use. Future work includes development and integration of new asset detection algorithms that could effectively recognize assets and localize their positions in 3D. More experiments also need to be conducted by expanding the training dataset, and testing the performance of the proposed method on different datasets with different levels of visibility. The 3D image-based reconstruction algorithm geo-registers images in 3D and as a result creates an opportunities that the outcome of any detection can be cross-correlated across multiple detections. The proposed methods as well as the overall pipeline lend themselves to applications in different contexts and with different imagery. For instance, the optimization for picking the best thresholds has a general formulation and parameters can be adapted based on training data. The number of annotations, the ratio between FPs and FNs, and the precision of segmentation can be used to set parameters.

5.1.2. Segmentation and Recognition of Roadway Assets from Car-Mounted Camera Video Streams using a Scalable Non-Parametric Image Parsing Method

This research also presented fast graph-based segmentation and super-parsing algorithms which efficiently segment roadway assets from 2D video streams. The state-of-the-art results on Smart Road and I-57 datasets were demonstrated. One of the main merits of the proposed method

is its reliance of lazy learning method. This framework does not need any supervised training except for on-time computation of basic statistics such as label co-occurrence probabilities. To achieve more comprehensive video frame understanding and to explore a higher-level form of context, we consider the task of simultaneously labeling regions into two types of classes: semantic and geometric. The local features through robust optimization of camera configuration provide acceptable performance thorough this simple concept. Taking the best scene matched from each of these global features leads to better superpixel-based matches for region-based features that capture similar types of cues as the global features.

5.1.3. Evaluation of Multi-Class Traffic Sign Detection and Classification Methods for U.S. Roadway Asset Inventory Management

Traffic sign recognition, particularly for different types is a challenging problem. In recent years, a lot of effort has gone into traffic sign recognition mainly from Europe, Japan, and Australia. Arguably, the most important issue with sign detection is currently lack of use of public image databases to train and test systems. Currently every new approach presented uses a new data set for testing which makes comparisons between different approaches hard. To facilitate more research in US traffic sign detection specifically, we contributed with a new database of nearly 11,000 signs. This work presented, validated, and compared video-based methods for detection and classification of multi-class of traffic signs which has potential to provide quick and inexpensive access to information about location and condition of traffic signs. It can also foster information sharing and exchange among different agencies as well as DOTs. Following the principle of sending a little time on the bulk of the data, and keeping a more refined analysis for the promising parts of the images, the proposed HOG+C system combines the efficiency with good performance. Most of the state-of-the-art traffic sign detectors rely on shape while ignoring color. A new method for traffic signs detection based using histograms of oriented gradients and Hue-Saturation colors was presented in this work. Color attributes are compact, computationally efficient, and possess some degree of photometric invariance while maintaining discriminative power.

5.1.4. Mapping Traffic Signs Using Google Street View Images for Roadway Inventory Management

By leveraging Google Street View images, this research presented a new system for creating comprehensive inventories of traffic signs. By processing images extracted using Google Street View API– using a computer vision method based on joint Histograms of Oriented Gradients and Color– traffic signs detected and classified into four categories of regulatory, warning, stop, and yield signs. Considering the discriminative classification scores from all images that see a sign, the most probable 3D location of each traffic sign was derived and shown on the Google Maps using a heat map. A data card containing information about location, type, and condition of each detected traffic sign was also created. Finally, several data mining interfaces were introduced that allow for better management of the traffic sign inventories. Given the reliability in performance shown through experiments and because collecting information from Google Street View imagery is cost-effective, the proposed method has potential to deliver inventory information on high-quantity low-cost assets in a timely fashion and tie into the existing DOT inventory management systems. In simple terms, the system outsources the task of data collection and in return provides an accurate geo-spatial localization of traffic signs along with useful information such as roadway number, city, state, zip-code, and type of traffic sign by visualizing them on the Google Map. It also provides automated inventory queries allowing professionals to spend less time searching for traffic signs, rather focus on the more important task of monitoring existing conditions.

5.1.5. Image-based Retro-Reflectivity Measurement of Traffic Signs in Day Time

One of the key measures for roadway safety at nighttime is sign visibility. Evaluation and replacing traffic signs with low retro-reflectivity is an effective strategy for improving safety of the transportation systems. With the new FHWA requirements on sign retro-reflectivity as outlined in MUTCD, road agencies require cost effective techniques that can enable retro-reflectivity measurement during daytime. To this end, this research presented and validated a new imaged-based method for measuring the retro-reflectivity of traffic signs in a daytime as a proof of concept. The method is an attempt toward measuring retro-reflectivity and has potential to provide quick, safe, and inexpensive compliance inspection for minimum level of retro-reflectivity in traffic signs. With the hardware mounted on a driving vehicle, these measurements can be taken remotely

and automatically for longer stretches of roadways and highways, and there is no further need for putting measurement equipment in contact with sign and repeating the manual process for one sign at a time.

5.2. Conclusion

Following a principle of spending little time on the bulk of the data, and keeping a more refined analysis for the promising parts of the images, the proposed system combines efficiency with good performance. The integer linear optimization formulation for selecting the optimal candidate extraction methods and the standard sliding window approach are found to be complementary to the proposed detection based on fast extracted candidates. The detected assets and their types are visualized in an augmented reality environment which enables remote walk-throughs for condition assessment purposes.

This research primarily segments video streams into different categories of assets, and does not fully recognize and classify different types of assets. Furthermore, it does not distinguish the intra-class variability in assets which is a key component in asset data collection and condition assessment; e.g., stop sign vs. speed limit sign. The proposed method does however enable DOT practitioners to identify 3D roadway assets. The results contribute to the body of knowledge by enabling for subsequent work on development of techniques that can further recognize the type of assets as a potential way to reduce the time and effort required for developing such inventories. The proposed system has potential to minimize the need for detection and identifying asset in each frames, or mapping them to previous assessments as conducted in previous section using Semantic Texton Forest and 3D point cloud model. This in turn has potential to allow the experts to only focus on the more important task of condition assessment and devising strategies on how prioritizations and improvements to existing conditions can be made. The new implementation based on the superpixels and lazy training algorithm has significantly reduced the computation time and make it feasible to segment the video frames in a continuous fashion which are typical in case of roadway infrastructure assets. Our experiment on Smart Road dataset with Semantic Texton Forest (STF) method took about a week for both training and testing, and now the new method is implemented just in 2-3days. The processing of test image in full resolution using the proposed method is approximately less than a minute which makes our method applicable for large datasets already collected by the DOTs. In particular, results on I-57 dataset which has 550 images

and 5,970 samples constitute an important benchmark for development and future approaches. To best of our knowledge, it is currently the most comprehensive dataset in high quantity low cost roadway asset. Finally, an extension of the proposed system to video segmentation and parsing is demonstrated. This extension segments the video into spatio-temporal supervoxels and uses a simple heuristic to combine local appearance cues across frames; however this method does not yet extract 3D geometry. Overall, the proposed method has two limitations: 1) the scene matching step for obtaining the retrieval set suffers from an inability of low-level global features such as Gist to retrieve semantically similar scenes, resulting in incoherent interpretations; 2) our reliance on bottom-up segmentation which affects the performance on roadway asset categories. Traditionally, such classes are handled using sliding window detectors to incorporate such detectors into region-based parsing. Nevertheless, our method can act as a fast and accurate basis for candidate selection leading to region-based parsing methods.

The results of multi-class traffic sign detection and classification with average performance accuracy of 94.83% shows the proposed approach can significantly improve the detection performance on the US traffic sign dataset and hold the promise of applicability for first step toward automated traffic sign detection and classification. We also evaluated other methods of Haar like features and HOG and effective parameters such as sliding window size on detection accuracy. The standard sliding window approach is found to be complementary for the proposed methods. The proposed multi-class sliding window is independent to scale and viewpoint of traffic signs, as well as illumination condition.

With the continuous growth and expansion of the roadway networks, the use of the proposed method will allow DOTs' practitioners to accommodate the demands of the installation of new traffic signs and other assets, maintain existing signs, and perform future replacements in compliance with the Manual on Uniform Traffic Control Devices (MUTCD). The report cards which contain latitude/longitude, roadway number, type of traffic sign, and detection/classification score facilitate the review of specific sign information in a given location without searching through the large databases. Such spatio-temporal representations provide DOTs with information on how different types of traffic signs degrade over time and further provides useful condition information necessary for predicting sign replacement plan.

Retro-reflectivity condition measurements can also be taken from real roadway geometries rather than prescribed geometries which do not always represent the real world conditions. In particular, the retro-reflectivity of twisted and leaning signs can also be measured under real roadway conditions and for actual driver-view perspectives. We evaluated our method with ground truth and showed that the proposed image-based method with current hardware setting is robust enough to measure retro-reflectivity of signs at different times of the day and for any distance less than 100*ft*. Such a mobile setup can significantly facilitate the current process by allowing inspection vehicles – widely used in the U.S. – to also measure retro-reflectivity levels during daytime. This methods can also minimizes the challenges associated with inspecting overhead and difficult-to-reach ground mounted signs.

5.3. Open Gaps-in-Knowledge

While this research presented the initial steps towards processing site video streams for the purpose of roadway asset categorization, several critical challenges remain. Some of the open research problems include:

- ***Segmentation and reconstruction***

  The method introduced in this dissertation primarily segments a point cloud into different categories of assets and does not distinguish the intra-class variability in assets, which is a key component in asset data collection and condition assessment. In the proposed method, the reconstructed 3D points, the 2D pixels, and their feature descriptors are all interlinked, enabling the future work to focus on improving the performance of recognition algorithms with the outcome of 3D reconstruction and segmentation; i.e. using geometry priors to improve recognition.

- ***3D localization of traffic signs and other high-quantity low-cost roadway assets in a large scale point cloud models***

  Using the method proposed in this dissertation, traffic signs can be detected and classify in 2D image and the user can localize assets in a supervised fashion. Once the 3D point cloud of roadways is available, the practitioners should select certain areas from 3D or their corresponding 2D regions to extract the location of assets in 3D environment. More works needs to be done on integrating asset detection algorithms with the presented work for

automated localization purposes. This asset 3D localization can be done by using connectivity semantics embedded between the video frames and 3D points in the reconstructed point cloud. This step needs to find at least one location per asset (ideally close to the center of the projection), while minimizing false alarms caused by observation from multiple video frames.

- ***Detection and classification of all types of traffic signs***

  In this research, the traffic signs were classified based on the signs' messages. For comprehensiveness, different warning and different regulatory signs with different pictogram and text were used as part of the training and testing datasets. There are more than 670 types of traffic signs specified in MUTCD and developing and validating the proposed system that can detect all types of traffic signs associated with MUTCD code is left as future work. For example, signs such as curve warning sign (W1-2) and road narrow warning sign (W5-1) are both in the dataset and can be classified them into the introduced four categories, however differencing between them is left as future work. The recent LISA dataset can also be fused with the introduced dataset for future experiments.

- ***Testing the proposed methods on local streets and non-interstate highways***

  Since there are no Stop Signs and very limited Yield Signs on interstate highways, the validation of our proposed methods for urban area is left as future work. Google Street View images can be an excellent source for this, yet more work needs to be done to test the performance of the new method on local streets and non-interstate highways.

- ***Detection and classification of traffic signs using mobile cameras***

  The ability to detect and classify traffic signs from moving cameras and commodity smartphones opens a great opportunity for developing autonomous vehicles. For example a consumer-level camera mounted on a car can help the development of autonomous vehicles and improve the safety. This can significantly cut down the cost of current efforts (e.g. Google autonomous vehicles) which use laser scanners. Using image-based localization methods on commodity smartphones to localize a field personnel to the integrated 3D model can streamline current inspections that still require manual input from

the users. Particularly it allows for user inputs to also be incorporated into the integrated model. Understanding and synthesizing information requirement, and developing methods for commodity smartphones is left as future work.

- ***Mounting proposed retro-reflectivity measurement system over an inspection vehicle***
  All retro-reflectivity measurements in our work were taken using a fixed setup. The presented method is an attempt for remotely images-based retro-reflectivity measurement of traffic signs in daytime and the concept of method was proofed. Hence, the impact of mounting our system over an inspection vehicle and the possible effects of vibrations and other factors in measuring retro-reflectivity for a longer stretch of road is not studied yet.

- ***Fully automated detection, classification, localization, and retro-reflectivity measurement of traffic signs***
  Using the proposed solutions of multi-class traffic sign detection and retro-reflectivity measurements has potential to automatically detect, classify, localize, and measure the retro-reflectivity of traffic signs in daytime; nevertheless this subject is left as future work.

REFERENCES

(FHWA), F. H. A. (2003). "Manual on uniform traffic control devices (MUTCD)." U.S. Department of Transportation, Washington, D.C.

Agrawal, A., Raskar, R., Nayar, S. K., and Li, Y. "Removing photography artifacts using gradient projection and flash-exposure sampling." *Proc., ACM Transactions on Graphics (TOG)*, ACM, 828-835.

Ai, C., and Tsai, Y. (2014). "Geometry Preserving Active Polygon-Incorporated Sign Detection Algorithm." *Journal of Computing in Civil Engineering*.

Ai, C., and Tsai, Y. J. (2011). "Hybrid Active Contour–Incorporated Sign Detection Algorithm." *Journal of Computing in Civil Engineering*, 26(1), 28-36.

Alefs, B., Eschemann, G., Ramoser, H., and Beleznai, C. "Road Sign Detection from Edge Orientation Histograms." *Proc., Intelligent Vehicles Symposium, 2007 IEEE*, 993-998.

Ali, N., Sobran, N., Ghazaly, M. M., Shukor, S. A., and Ibrahim, A. F. T. (2014). "Traffic Sign Detection and Classification for Driver Assistant System." *The 8th International Conference on Robotic, Vision, Signal Processing & Power Applications*, H. A. Mat Sakim, and M. T. Mustaffa, eds., Springer Singapore, 277-283.

ASCE, A. S. o. C. E. (2013). "National Infrastructure Report Card."

Ashouri Rad, A., and Rahmandad, H. (2013). "Reconstructing Online Behaviors by Effort Minimization." *Social Computing, Behavioral-Cultural Modeling and Prediction*, A. Greenberg, W. Kennedy, and N. Bos, eds., Springer Berlin Heidelberg, 75-82.

Austin, R. L., Schultz, R. J., and Scientific, G. (2009). *Guide to Retroreflection Safety Principles and Retrorereflective [sic] Measurements*, Gamma Scientific.

Babić, D., Ščukanec, A., and Fiolić, M. "Traffic sign analysis as a function of traffic safety on Croatian state road D3." *Proc., ICTTE-INTERNATIONAL CONFERENCE ON TRAFFIC AND TRANSPORT ENGINEERING*.

Bahlmann, C., Ying, Z., Ramesh, V., Pellkofer, M., and Koehler, T. "A system for traffic sign detection, tracking, and recognition using color, shape, and motion information." *Proc., Intelligent Vehicles Symposium, 2005. Proceedings. IEEE*, 255-260.

Balali, V., Ashouri Rad, A., and Golparvar-Fard, M. (2015). "Detection, Classification, and Mapping of U.S. Traffic Signs Using Google Street View Images for Roadway Inventory Management." *Journal of Transportation Part C: Emerging Technologies*.

Balali, V., Depwe, E., and Golparvar-Fard, M. "Multi-class Traffic Sign Detection and Classification Using Google Street View Images." *Proc., Transportation Research Board 94th Annual Meeting*, TRB.

Balali, V., and Golparvar-Fard, M. (2014). "Scalable Non-Parametric Parsing for Segmentation and 3D Localization of High-Quantity Low-Cost Highway Assets from Car-Mounted Video Streams." *Construction Research Congress (CRC)*Atlanta, GA, USA, 120-129.

Balali, V., and Golparvar-Fard, M. (2014). "Segmentation and Recognition of Roadway Assets from Car-Mounted Camera Video Streams using a Scalable Non-Parametric Image Parsing Method." *Automation in Construction*.

Balali, V., and Golparvar-Fard, M. (2014). "Video-Based Detection and Classification of US Traffic Signs and Mile Markers using Color Candidate Extraction and Feature-Based Recognition." *Computing in Civil and Building Engineering*, 858-866.

Balali, V., and Golparvar-Fard, M. (2015). "Evaluation of Multi-Class Traffic Sign Detection and Classification Methods for U.S. Roadway Asset Inventory Management." *ASCE Journal of Computing in Civil Engineering*.

Balali, V., and Golparvar-Fard, M. (2015). "Segmentation and recognition of roadway assets from car-mounted camera video streams using a scalable non-parametric image parsing method." *Automation in Construction*, 49, 27-39.

Balali, V., Golparvar-Fard, M., and de la Garza, J. M. "Video-based Highway Asset Recognition and 3D Localization." *Proc., Computing in Civil Engineering*, 379-386.

Balali, V., Muthukumar, B., and Golparvar-Fard, M. (2015). "Recognition and Localization of Traffic Signs in Google Maps via 3D Image-based Point Cloud Models." *ASCE International Workshop on Computing in Civil Engineering*, ASCE, Austin, TX.

Ballerini, R., Cinque, L., Lombardi, L., and Marmo, R. (2005). "Rectangular Traffic Sign Recognition." *Image Analysis and Processing – ICIAP 2005*, F. Roli, and S. Vitulano, eds., Springer Berlin Heidelberg, 1101-1108.

Baro, X., Escalera, S., Vitria, J., Pujol, O., and Radeva, P. (2009). "Traffic Sign Recognition Using Evolutionary Adaboost Detection and Forest-ECOC Classification." *Intelligent Transportation Systems, IEEE Transactions on*, 10(1), 113-126.

Beshah, T., and Hill, S. "Mining Road Traffic Accident Data to Improve Safety: Role of Road-Related Factors on Accident Severity in Ethiopia." *Proc., AAAI Spring Symposium: Artificial Intelligence for Development*.

Bhalla, S., Naidu, A. S., and Soh, C. K. "Influence of structure-actuator interactions and temperature on piezoelectric mechatronic signatures for NDE." *Proc., Smart Materials, Structures, and Systems*, International Society for Optics and Photonics, 263-269.

Bianchini, A., Bandini, P., and Smith, D. W. (2010). "Interrater Reliability of Manual Pavement Distress Evaluations." *Journal of Transportation Engineering*, 136(2), 165-172.

Brilakis, I., Fathi, H., and Rashidi, A. (2011). "Progressive 3D reconstruction of infrastructure with videogrammetry." *Automation in Construction*, 20(7), 884-895.

Brimley, B., and Ye, F. "Measurement Bias and Reproducibility of In-Service Sign Retroreflectivity Readings Made with Handheld Instruments." *Proc., Transportation Research Board 92nd Annual Meeting*.

Brkic, K. (2010). "An overview of traffic sign detection methods." *Department of Electronics, Microelectronics, Computer and Intelligent Systems Faculty of Electrical Engineering and Computing Unska*, 3, 10000.

Brkic, K. (2013). "An overview of traffic sign detection methods." Department of Electronics, Microelectronics, Computer and Intelligent Systems, Faculty of Electrical Engineering and Computing, Unska, Zagreb, Croatia, 10000.

Brkic, K. (2013). "An overview of traffic sign detection methods.", Unska, Zagreb, Croatia., Faculty of Electrical Engineering and Computing.

Brostow, G. J., Shotton, J., Fauqueur, J., and Cipolla, R. (2008). "Segmentation and Recognition Using Structure from Motion Point Clouds." *Proc., the 10th European Conference on Computer Vision: Part I*, Springer-Verlag, Marseille, France, 44-57.

Burges, C. C. (1998). "A Tutorial on Support Vector Machines for Pattern Recognition." *Data Mining and Knowledge Discovery*, 2(2), 121-167.

Caddell, R., Hammond, P., and Reinmuth, S. (2009). "Roadside Features Inventory Program." *Washington State Department of Transportation*.

Carlson, P. J., and Lupes, M. S. (2007). "Methods for maintaining traffic sign retroreflectivity."

Carlson, P. J., and Picha, D. (2009). "Sign Retroreflectivity Guidebook.", Federal Highway Administartion.
Chang, L.-Y., and Chen, W.-C. (2005). "Data mining of tree-based models to analyze freeway accident frequency." *Journal of Safety Research*, 36(4), 365-375.
Chen, H., and Wolff, L. B. (1998). "Polarization phase-based method for material classification in computer vision." *International Journal of Computer Vision*, 28(1), 73-83.
Cheok, G., Franaszek, M., Katz, I., Lytle, A., Saidi, K., and Scott, N. (2010). "Assessing Technology Gaps for the Federal Highway Administration Digital Highway Measurement Program.", Construction Metrology and Automation Group, Building and Fire Research Laboratory, National Institute of Standards and Technology (NIST).
Cimpoi, M. (2011). "Traffic sign detection and classification in video mode." *eRAF Journal on Computing*, 3, 10-16.
Conrad, J. (1998). "Exposure Metering."
Crandall, D., Owens, A., Snavely, N., and Huttenlocher, D. (2011). "Discrete-continuous optimization for large-scale structure from motion." *Conference on Computer Vision and Pattern Recognition (CVPR)*, 3001-3008.
Creusen, I., and Hazelhoff, L. "A semi-automatic traffic sign detection, classification, and positioning system." *Proc., IS&T/SPIE Electronic Imaging*, International Society for Optics and Photonics, 83050Y-83050Y-83056.
Creusen, I. M., Wijnhoven, R. G. J., Herbschleb, E., and de With, P. H. N. "Color exploitation in hog-based traffic sign detection." *Proc., Image Processing (ICIP), 2010 17th IEEE International Conference on*, 2669-2672.
CTRE (2004). "Digital Satellite Images." <http://www.ctre.iastate.edu/research/bts_wb/cd-rom/spatial/dsi.htm>.
Dalal, N., and Triggs, B. "Histograms of oriented gradients for human detection." *Proc., Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, IEEE, 886-893.
DARPA (2012). "Defense Advanced Research Projects Agency."
De la Escalera, A., Armingol, J. M., and Mata, M. (2003). "Traffic sign recognition and analysis for intelligent vehicles." *Image and vision computing*, 21(3), 247-258.
de la Garza, J., Howerton, C., and Sideris, D. "A study of implementation of IP-S2 mobile mapping technology for highway asset condition assessment." *Proc., Computing in Civil Engineering (2011)*, ASCE, 1-8.
de la Garza, J., Howerton, C., and Sideris, D. (2011). "A Study of Implementation of IP-S2 Mobile Mapping Technology for Highway Asset Condition Assessment." *Computing in Civil Engineering (2011)*, 1-8.
de la Garza, J., Roca, I., and Sparrow, J. "Visualization of failed highway assets through geocoded pictures in google earth and google maps." *Proc., Proc., CIB W078 27th International Conference on Applications of IT in the AEC Industry*.
de la Garza, J., Yates, N., and Arrington, A. "Improving highway asset management with RFID technology." *Proc., Proceedings of the 26th CIBW078 Conference in Istanbul, Turkey*, 163-169.
de la Garza, J. M., Roca, I., and Sparrow, J. "Visualization of failed highway assets through geocoded pictures in google earth and google maps." *Proc., Proc., the CIB W078 27th International Conference on Applications of IT in the AEC Industry, .*

de la Garza, J. M., Roca, I., and Sparrow, J. (2010). "Visualization of failed highway assets through geocoded pictures in google earth and google maps." *Proc., the CIB W078 27th International Conference on Applications of IT in the AEC Industry*, Cairo, Egypt.

de la Garza, J. M., Yates, N. J., and Arrington, C. A. (2010). "Improving highway asset management with RFID technology." *Managing IT in Construction/Managing Construction for Tomorrow*London, UK.

de la Torre, J. "Organising geo-temporal data with CartoDB, an open source database on the cloud." *Proc., Biodiversity Informatics Horizons 2013*.

Debevec, P. E., and Malik, J. "Recovering high dynamic range radiance maps from photographs." *Proc., ACM SIGGRAPH 2008 classes*, ACM, 31.

DeGray, J., and Hancock, K. L. (2002). "Ground-based image and data acquisition systems for roadway inventories in New England: A synthesis of highway practice." New England Transportation Consortium.

Evans, T., Heaslip, K., Boggs, W., Hurwitz, D., and Gardiner, K. (2012). "Assessment of sign retroreflectivity compliance for development of a management plan." *Transportation Research Record: Journal of the Transportation Research Board*, 2272(1), 103-112.

Fang, C.-Y., Chen, S.-W., and Fuh, C.-S. (2003). "Road-sign detection and tracking." *Vehicular Technology, IEEE Transactions on*, 52(5), 1329-1341.

Fatmehsan, Y., Ghahari, A., and Zoroofi, R. "Gabor wavelet for road sign detection and recognition using a hybrid classifier." *Proc., Multimedia Computing and Information Technology (MCIT), 2010 International Conference on*, IEEE, 25-28.

FeiXiang, R., Jinsheng, H., Ruyi, J., and Klette, R. "General traffic sign recognition by feature matching." *Proc., Image and Vision Computing New Zealand, 2009. IVCNZ '09. 24th International Conference*, 409-414.

Felzenszwalb, P. F., Girshick, R. B., McAllester, D., and Ramanan, D. (2010). "Object detection with discriminatively trained part-based models." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(9), 1627-1645.

Felzenszwalb, P. F., and Huttenlocher, D. P. (2004). "Efficient graph-based image segmentation." *International Journal of Computer Vision*, 59(2), 167-181.

FHWA, F. H. A. (2005). "Asset Management Systems for Roadway Safety." Federal Highway Administration.

FHWA, F. H. A. (2009). "Manual on uniform traffic control devices (MUTCD)." U.S. Department of Transportation,, Washington, D.C.

FHWA, F. H. A. (2010). "FHWA-HRT-05-077."

FHWA, F. H. A. (2010). "Highway Safety and Asset Management."

Fischler, M. A., and Bolles, R. C. (1981). "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography." *Communications of the ACM*, 24(6), 381-395.

Flintsch, G. W., and Bryant, J. W. (2009). "Asset Management Data Collection for Supporting Decision Processes." Federal Highway Administration (FHWA).

Frahm, J. M., Pollefeys, M., Lazebnik, S., Gallup, D., Clipp, B., Raguram, R., Wu, C., Zach, C., and Johnson, T. (2010). "Fast Robust Large-scale Mapping from Video and Internet Photo Collections." *In special issue "100 years of ISPRS" of the ISPRS Journal of Photogrammetry and Remote Sensing*, 65(6).

Furukawa, Y., Curless, B., Seitz, S. M., and Szeliski, R. "Reconstructing building interiors from images." *Proc., Computer Vision, 2009 IEEE 12th International Conference on*, IEEE, 80-87.

Furukawa, Y., Curless, B., Seitz, S. M., and Szeliski, R. (2010). "Towards Internet-scale multi-view stereo." *Proc., Computer Vision and Pattern Recognition*, 1434-1441.

Galleguillos, C., and Belongie, S. (2010). "Context based object categorization: A critical survey." *Computer Vision and Image Understanding*, 114(6), 712-722.

Gallup, D., Frahm, J. M., and Pollefeys, M. (2010). "A Heightmap Model for Efficient 3D Reconstruction from Street-Level Video." *Proc., International Conference on 3D Data Processing, Visualization and Transmission.*

Gao, X. W., Podladchikova, L., and Shaposhnikov, D. (2003). "Application of Vision Models to Traffic Sign Recognition." *Artificial Neural Networks and Neural Information Processing — ICANN/ICONIP 2003*, O. Kaynak, E. Alpaydin, E. Oja, and L. Xu, eds., Springer Berlin Heidelberg, 1100-1105.

Gao, X. W., Podladchikova, L., Shaposhnikov, D., Hong, K., and Shevtsova, N. (2006). "Recognition of traffic signs based on their colour and shape features extracted using human vision models." *Journal of Visual Communication and Image Representation*, 17(4), 675-685.

Geronimo, D., Lopez, A. M., Sappa, A. D., and Graf, T. (2010). "Survey of Pedestrian Detection for Advanced Driver Assistance Systems." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 32(7), 1239-1258.

Gil-Jim, P., #233, nez, Lafuente-Arroyo, S., Maldonado-Basc, S., #243, G, H., #243, and mez-Moreno (2005). "Shape classification algorithm using support vector machines for traffic sign recognition." *Proceedings of the 8th international conference on Artificial Neural Networks: computational Intelligence and Bioinspired Systems*, Springer-Verlag, Barcelona, Spain, 873-880.

Golparvar-Fard, M., Balali, V., and de la Garza, J. M. (2012). "Segmentation and recognition of highway assets using image-based 3D point clouds and semantic Texton forests." *Journal of Computing in Civil Engineering*(04014023).

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2009). "Sparse reconstruction and geo-registration of daily site photographs for representation of as-built construction scene and automatic construction progress data collection." *Proc., International Symposium on Automation and Robotics in Construction*Austin, TX, USA.

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2010). "D4AR – 4 Dimensional augmented reality - tools for automated remote progress tracking and support of decision-enabling tasks in the AEC/FM industry." *Proc., the 6th Int. Conf. on Innovations in AEC.*

Golparvar-Fard, M., Peña-Mora, F., and Savarese, S. (2012b). "Automated operation-level tracking of progress using unordered daily construction photographs and IFC as-planned models." *Journal of Computing in Civil Engineering(ASCE), In Press.*

Gomez-Moreno, H., Maldonado-Bascon, S., Gil-Jimenez, P., and Lafuente-Arroyo, S. (2010). "Goal Evaluation of Segmentation Algorithms for Traffic Sign Recognition." *Intelligent Transportation Systems, IEEE Transactions on*, 11(4), 917-930.

Gong, J., Zhou, H., Gordon, C., and Jalayer, M. (2012). "Mobile Terrestrial Laser Scanning for Highway Inventory Data Collection." *Computing in Civil Engineering*, 545-552.

Gong, J., Zhou, H., Gordon, C., and Jalayer, M. "Mobile terrestrial laser scanning for highway inventory data collection." *Proc., Proceedings of International Conference on Computing in Civil Engineering, Clearwater Beach, FL, USA*, 17-20.

Gonzalez, H., Halevy, A. Y., Jensen, C. S., Langen, A., Madhavan, J., Shapley, R., Shen, W., and Goldberg-Kidon, J. "Google fusion tables: web-centered data management and collaboration." *Proc., Proceedings of the 2010 ACM SIGMOD International Conference on Management of data*, ACM, 1061-1066.

Gould, S., Fulton, R., and Koller, D. "Decomposing a scene into geometric and semantically consistent regions." *Proc., Computer Vision, 2009 IEEE 12th International Conference on*, IEEE, 1-8.

Grundmann, M., Kwatra, V., Han, M., and Essa, I. "Efficient hierarchical graph-based video segmentation." *Proc., Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, IEEE, 2141-2148.

Guizar-Sicairos, M., Thurman, S. T., and Fienup, J. R. (2008). "Efficient subpixel image registration algorithms." *Optics letters*, 33(2), 156-158.

Haas, K., and Hensing, D. (2005). "Why your agency should consider asset management systems for roadway safety."

Haas, K., and Hensing, D. (2005). "Why your agency should consider asset management systems for roadway safety."

Harris, E. A. (2007). "SIGN MAINTENANCE STRATEGIES FOR AGENCIES TO COMPLY WITH PROPOSED FHWA MINIMUM RETROREFLECTIVITY STANDARDS." North Carolina State University.

Hartley, R., and Zisserman, A. (2003). *Multiple view geometry in computer vision*, Cambridge university press.

Hauser, T. A., and Scherer, W. T. "Data mining tools for real-time traffic signal decision support & maintenance." *Proc., Systems, Man, and Cybernetics, 2001 IEEE International Conference on*, 1471-1477 vol.1473.

Heng, L., Lee, G. H., Fraundorfer, F., and Pollefeys, M. (2011). "Real-time photo-realistic 3D mapping for micro aerial vehicles." *Proc., IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 4012-4019.

Hiscocks, P. D., and Eng, P. (2011). "Measuring luminance with a digital camera." Syscomp electronic design limited.

Horn, B. K., and Schunck, B. G. "Determining optical flow." *Proc., 1981 Technical Symposium East*, International Society for Optics and Photonics, 319-331.

Houben, S. "A single target voting scheme for traffic sign detection." *Proc., Intelligent Vehicles Symposium (IV), 2011 IEEE*, 124-129.

Hsin-Han, C., Yen-Lin, C., Wen-Qing, W., and Tsu-Tian, L. "Road speed sign recognition using edge-voting principle and learning vector quantization network." *Proc., Computer Symposium (ICS), 2010 International*, 246-251.

Hu, X., Tao, C. V., and Hu, Y. (2004). "Automatic road extraction from dense urban area by integrated processing of high resolution imagery and lidar data." *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Istanbul, Turkey*, 35, B3.

Hu, Z., and Tsai, Y. (2011). "Generalized image recognition algorithm for sign inventory." *Journal of Computing in Civil Engineering*, 25(2), 149-158.

Hu, Z., and Tsai, Y. (2011b). "Image Recognition Model for Developing a Sign Inventory." *ASCE Journal of Computing in Civil Engineering*, 25(2), 385-400.
Hulme, E., Hubbard, S. M., Farnsworth, G. D., Hainen, A. M., Remias, S. M., and Bullock, D. M. (2011). "An Asset Management Framework for Addressing the New MUTCD Traffic Sign Retroreflectivity Standards."
Hummer, J. E., Harris, E. A., and Rasdorf, W. (2013). "Simulation-Based Evaluation of Traffic Sign Retroreflectivity Maintenance Practices." *Journal of Transportation Engineering*, 139(6), 556-564.
Jalayer, M., Gong, J., Zhou, H., and Grinter, M. "Evaluation of Remote-Sensing Technologies for Collecting Roadside Feature Data to Support Highway Safety Manual Implementation." *Proc., Transportation Research Board 92nd Annual Meeting*.
Jaselskis, E. J., Cackler, E. T., Walters, R. C., Zhang, J., and Kaewmoracharoen, M. (2006). "Using scanning lasers for real-time pavement thickness measurement."
Jeyapalan, K. (2004). "Mobile digital cameras for as-built surveys of roadside features." *Photogrammetric Engineering & Remote Sensing*, 70(3), 301-312.
Jeyapalan, K., and Jaselskis, E. (2002). "Technology Transfer of As-Built and Preliminary Surveys Using GPS, Soft Photogrammetry, and Video Logging."
Jin, Y., Dai, J., and Lu, C.-T. "Spatial-temporal data mining in traffic incident detection." *Proc., Proc. SIAM DM 2006 Workshop on Spatial Data Mining*, Citeseer.
Johnson, M., and Shotton, J. (2010). "Semantic Texton Forests." *Computer Vision*, R. Cipolla, S. Battiato, and G. Farinella, eds., Springer Berlin Heidelberg, 173-203.
Johnson, S. (2006). *Stephen Johnson on digital photography*, O'Reilly Media, Inc.
Jones, F. E. (2004). "GPS-based Sign Inventory and Inspection Program." *International Municipal Signal Association (IMSA) Journal*, 30-35.
Kapler, T., and Wright, W. (2005). "GeoTime information visualization." *Information Visualization*, 4(2), 136-146.
Kavulya, G., Jazizadeh, F., and Becerik-Gerber, B. (2011). "Effects of Color, Distance, And Incident angle on Quality of 3D point clouds." *Computing in Civil Engineering*, 21-21.
Keller, C. G., Sprunk, C., Bahlmann, C., Giebel, J., and Baratoff, G. "Real-time recognition of U.S. speed signs." *Proc., Intelligent Vehicles Symposium, 2008 IEEE*, 518-523.
Khalilikhah, M., Heaslip, K., and Louisell, C. (2015). "Analysis of the Effects of Coarse Particulate Matter (PM10) on Traffic Sign Retroreflectivity." *Transportation Research Board 94th Annual Meeting*, TRB, Washington DC.
Khattak, A. J., HUMMER, J. E., and KARIMI, H. A. (2000). "New and existing roadway inventory data acquisition methods." *Journal of Transportation and Statistics*, 33.
Kianfar, J., and Edara, P. (2013). "A Data Mining Approach to Creating Fundamental Traffic Flow Diagram." *Procedia - Social and Behavioral Sciences*, 104(0), 430-439.
Kim, J., Jung, K., and Hyun, C. (2005). "A Study on an Efficient Sign Recognition Algorithm for a Ubiquitous Traffic System on DSP." *Computational Science and Its Applications – ICCSA 2005*, O. Gervasi, M. Gavrilova, V. Kumar, A. Laganà, H. Lee, Y. Mun, D. Taniar, and C. Tan, eds., Springer Berlin Heidelberg, 1177-1186.
Kiziltas, S., Burcu, A., Ergen, E., and Pingbo, T. (2008). "Technological assessment and process implications of field data capture technologies for construction and facility/infrastructure management." ITcon.
Krishnan, A. (2009). "Computer vision system for identifying road signs using triangulation and bundle adjustment." M.Sc., Kansas State University, Manhattan, Kansas.

Ladick, L., Sturgess, P., Alahari, K., Russell, C., and Torr, P. H. S. "What, where and how many? combining object detectors and CRFs." *Proc., 11th European conference on Computer vision: Part IV*, Springer-Verlag, 1888122, 424-437.

Larson, C. D., and Skrypczuk, O. (2004). "Comprehensive data collection to support asset management at Virginia Department of Transportation." *Transportation Research Record: Journal of the Transportation Research Board*, 1885(1), 96-103.

Larsson, F., and Felsberg, M. (2011). "Using fourier descriptors and spatial models for traffic sign recognition." *Proceedings of the 17th Scandinavian conference on Image analysis*, Springer-Verlag, Ystad, Sweden, 238-249.

Lazaros, N., Sirakoulis, G. C., and Gasteratos, A. (2008). "Review of stereo vision algorithms: from software to hardware." *International Journal of Optomechatronics*, 2(4), 435-462.

Lazebnik, S., Schmid, C., and Ponce, J. "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories." *Proc., Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*, IEEE, 2169-2178.

Lepetit, V., Lagger, P., and Fua, P. "Randomized trees for real-time keypoint recognition." *Proc., Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, IEEE, 775-781.

Li, D., and Su, W. Y. (2014). "Dynamic Maintenance Data Mining of Traffic Sign Based on Mobile Mapping System." *Applied Mechanics and Materials*, 455, 438-441.

Li, Z. (2008). "Project 08-06 June 2008." Illinois Institute of Technology.

Liu, C., Yuen, J., and Torralba, A. (2011). "Nonparametric scene parsing via label transfer." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(12), 2368-2382.

Lopez, L. D., and Fuentes, O. (2007). "Color-based road sign detection and tracking." *Proceedings of the 4th international conference on Image Analysis and Recognition*, Springer-Verlag, Montreal, Canada, 1138-1147.

Loy, G., and Barnes, N. "Fast shape-based road sign detection for a driver assistance system." *Proc., Intelligent Robots and Systems, 2004. (IROS 2004). Proceedings. 2004 IEEE/RSJ International Conference on*, 70-75 vol.71.

Maerz, N. H., and McKenna, S. (1999). "Mobile highway inventory and measurement system." *Transportation Research Record: Journal of the Transportation Research Board*, 1690(1), 135-142.

Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H., and Lopez-Ferreras, F. (2007). "Road-Sign Detection and Recognition Based on Support Vector Machines." *Intelligent Transportation Systems, IEEE Transactions on*, 8(2), 264-278.

Maldonado Bascón, S., Acevedo Rodríguez, J., Lafuente Arroyo, S., Fernndez Caballero, A., and López-Ferreras, F. (2010). "An optimization on pictogram identification for the road-sign recognition task using SVMs." *Computer Vision and Image Understanding*, 114(3), 373-383.

Malisiewicz, T., and Efros, A. A. "Recognition by association via learning per-exemplar distances." *Proc., Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, IEEE, 1-8.

Markow, M. J. (2007). *Managing selected transportation assets: Signals, lighting, signs, pavement markings, culverts, and sidewalks*, Transportation Research Board.

Mashford, J., Davis, P., and Rahilly, M. (2007). "Pixel-Based Colour Image Segmentation Using Support Vector Machine for Automatic Pipe Inspection." *AI 2007: Advances in Artificial Intelligence*, M. Orgun, and J. Thornton, eds., Springer Berlin Heidelberg, 739-743.

Mathias, M., Timofte, R., Benenson, R., and Van Gool, L. "Traffic sign recognition: How far are we from the solution?" *Proc., Neural Networks (IJCNN), The 2013 International Joint Conference on*, 1-8.

Medina, R. A., Haghani, A., and Harris, N. (2009). "Sampling protocol for condition assessment of selected assets." *Journal of Transportation Engineering*, 135(4), 183-196.

Meegoda, J. N., Juliano, T. M., and Banerjee, A. (2006). "Framework for Automatic condition Assessment of Culverts." *Transportation Research Record: Journal of the Transportation Research Board*, 1948, 26-34.

Meuter, M., Kummert, A., and Muller-Schneiders, S. "3d traffic sign tracking using a particle filter." *Proc., Intelligent Transportation Systems, 2008. ITSC 2008. 11th International IEEE Conference on*, IEEE, 168-173.

Miura, J., Kanda, T., and Shirai, Y. "An active vision system for real-time traffic sign recognition." *Proc., Intelligent Transportation Systems, 2000. Proceedings. 2000 IEEE*, IEEE, 52-57.

Mizusawa, D., and McNeil, S. (2006). "The Role of Advanced Technology in Asset Management: International Experiences." *Applications of Advanced Technology in Transportation 2006*.

MNDOT, M. D. o. T. (2009). "Pavement Condition Executive Summary." Minnesota Department of Transportation

Mogelmose, A., Trivedi, M. M., and Moeslund, T. B. (2012). "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: Perspectives and survey." *Intelligent Transportation Systems, IEEE Transactions on*, 13(4), 1484-1497.

Mordohai, P., Frahm, J.-M., Akbarzadeh, A., Clipp, B., Engels, C., Gallup, D., Merrell, P., Salmi, C., Sinha, S., and Talton, B. (2007). "Real-time video-based reconstruction of urban environments." *ISPRS Working Group*, 4.

Mordohai, P., Frahm, J. M., Akbarzadeh, A., Clipp, B., Engels, C., Gallup, D., Merrell, P., Salmi, C., Sinha, S., Talton, B., Wang, L., Yang, Q., Stewenius, H., Towles, H., Welch, G., Yang, R., Pollefeys, M., and Nister, D. (2007). "Real-Time Video-Based Reconstruction of Urban Environments." *Proc., 3DARCH: 3D Virtual Reconstruction and Visualization of Complex Architectures*.

NAE, N. A. o. E. (2010). "Grand Challenges for Engineering. NAE of the National Academies.", NAE, National Academy of Engineers.

Nakata, T., and Takeuchi, J.-i. "Mining traffic data from probe-car system for travel time prediction." *Proc., Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, ACM, 817-822.

NASA (2000). "Remote Sensing for Transportation.", Washington, DC.

NCRST (2001). "Remote Sensing and Spatial Information Technologies in Transportation." Synthesis Report, National Consortium on Remote Sensing in Transportation, University of California, Santa Barbara, CA.

Nistér, D. (2004). "An efficient solution to the five-point relative pose problem." *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 26(6), 756-770.

Oliva, A., and Torralba, A. (2006). "Building the gist of a scene: The role of global image features in recognition." *Progress in brain research*, 155, 23-36.

Overett, G., Tychsen-Smith, L., Petersson, L., Pettersson, N., and Andersson, L. (2011). "Creating robust high-throughput traffic sign detectors using centre-surround HOG statistics." *Machine Vision and Applications*, 1-14.

Pettersson, N., Petersson, L., and Andersson, L. "The histogram feature - a resource-efficient Weak Classifier." *Proc., Intelligent Vehicles Symposium, 2008 IEEE*, 678-683.

Pike, A., and Carlson, P. (2013). "Investigate The Ability To Determine Pavement Marking Retroreflectivity - Evaluate the Advanced Mobile Asset Collection (AMAC) Mobile Pavement Marking Retroreflectivity Measurement System." TEXAS A&M TRANSPORTATION INSTITUTE, College Station, TX.

Preston, H., Atkins, K. C., Lebens, M., and Jensen, M. (2014). "Traffic Sign Life Expectancy."

Prisacariu, V. A., Timofte, R., Zimmermann, K., Reid, I., and van Gool, L. "Integrating Object Detection with 3D Tracking Towards a Better Driver Assistance System." *Proc., Pattern Recognition (ICPR), 2010 20th International Conference on*, 3344-3347.

Rasdorf, W., Hummer, J. E., Harris, E. A., and Sitzabee, W. E. (2009). "IT Issues for the Management of High-Quantity, Low-Cost Assets." *Journal of Computing in Civil Engineering*, 23(2), 91-99.

Raskar, R., Tan, K.-H., Feris, R., Yu, J., and Turk, M. "Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging." *Proc., ACM Transactions on Graphics (TOG)*, ACM, 679-688.

Ravani, B., Dart, M., Hiremagalur, J., Lasky, T. A., and Tabib, S. (2009). "Inventory and Assessing Conditions of Roadside Features Statewide.", California State Department of Transportation., Advanced Highway Maintenance and Construction Technology Research Center.

Retterath, J. E., and Laumeyer, R. A. (2004). "System for automated determination of retroreflectivity of road signs and other reflective objects." Google Patents.

Reynolds, T. L. (2012). "Determining a Strategy for Efficiently Managing Sign Retroreflectivity in New Hampshire."

Robyak, R., and Orvets, G. (2004). "Video based Asset Data Collection at NJDOT.", New Jersey Department of Transportation.

Russell, B. C., Torralba, A., Murphy, K. P., and Freeman, W. T. (2008). "LabelMe: a database and web-based tool for image annotation." *International journal of computer vision*, 77(1-3), 157-173.

Ruta, A., Li, Y., and Liu, X. (2010). "Real-time traffic sign recognition from video by class-specific discriminative features." *Pattern Recognition*, 43(1), 416-430.

Scaramuzza, D., Fraundorfer, F., Pollefeys, M., and Siegwart, R. "Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints." *Proc., Computer Vision, 2009 IEEE 12th International Conference on*, IEEE, 1413-1419.

Scaramuzza, D., Fraundorfer, F., Pollefeys, M., and Siegwart, R. (2009). "Absolute scale in structure from motion from a single vehicle mounted camera by exploiting nonholonomic constraints." *12th International Conference on Computer Vision in 2009 IEEE*, 1413-1419.

Schechner, Y. Y., Narasimhan, S. G., and Nayar, S. K. (2003). "Polarization-based vision through haze." *Applied Optics*, 42(3), 511-525.

Shcukanec, A., Babic, D., and Sokol, H. (2014). "METHODOLOGY FOR MEASURING TRAFFIC SIGNS RETROREFLECTION." *European Scientific Journal*, 10(7).

Shotton, J., Johnson, M., and Cipolla, R. (2008). "Semantic texton forests for image categorization and segmentation." *Proc., Conference on Computer Vision and Pattern Recognition (CVPR 2008)*, 1-8.

Shotton, J., Winn, J., Rother, C., and Criminisi, A. (2009). "Textonboost for image understanding: Multi-class object recognition and segmentation by jointly modeling texture, layout, and context." *International Journal of Computer Vision*, 81(1), 2-23.

Shuang-dong, Z., Zhany, Y., and Lu, X.-f. "Detection for triangle traffic sign based on neural network." *Proc., Vehicular Electronics and Safety, 2005. IEEE International Conference on*, 25-28.

Smith, K., and Fletcher, A. (2001). "Evaluation of the FHWA's Sign Management and Retroreflectivity Tracking System (SMARTS) Van."

Snavely, N., Garg, R., Seitz, S. M., and Szeliski, R. (2008). "Finding Paths through the World's Photos." *Proc., ACM Transactions on Graphics (SIGGRAPH)*.

Soheilian, B., Paparoditis, N., and Vallet, B. (2013). "Detection and 3D reconstruction of traffic signs from multiple view color images." *ISPRS Journal of Photogrammetry and Remote Sensing*, 77(0), 1-20.

Stallkamp, J., Schlipsing, M., Salmen, J., and Igel, C. "The German traffic sign recognition benchmark: a multi-class classification competition." *Proc., Neural Networks (IJCNN), The 2011 International Joint Conference on*, IEEE, 1453-1460.

Stallkamp, J., Schlipsing, M., Salmen, J., and Igel, C. (2012). "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition." *Neural networks*, 32, 323-332.

Svennerberg, G. (2010). "Dealing with Massive Numbers of Markers." *Beginning Google Maps API 3*, M. Wade, C. Andres, S. Anglin, M. Beckner, E. Buckingham, G. Cornell, J. Gennick, J. Hassell, M. Lowman, M. Moodie, D. Parkes, J. Pepper, F. Pohlmann, D. Pundick, B. Renow-Clarke, D. Shakeshaft, T. Welsh, M. Tobin, J. Blackwell, and K. Wimpsett, eds., Apress, 177-210.

Tighe, J., and Lazebnik, S. (2010). "Superparsing: scalable nonparametric image parsing with superpixels." *Computer Vision–ECCV 2010*, Springer, 352-365.

Tighe, J., and Lazebnik, S. (2013). "Superparsing- Scalable Nonparametric Image Parsing with Superpixels." *International Journal of Computer Vision*, 101(2), 329-349.

Timofte, R., Zimmermann, K., and Luc Van, G. "Multi-view traffic sign detection, recognition, and 3D localisation." *Proc., Applications of Computer Vision (WACV), 2009 Workshop on*, 1-8.

Timofte, R., Zimmermann, K., and Van Gool, L. (2014). "Multi-view traffic sign detection, recognition, and 3D localisation." *Machine Vision and Applications*, 25(3), 633-647.

Torrent, D. G., and Caldas, C. H. (2009). "Methodology for automating the identification and localization of construction components on industrial projects." *Journal of Computing in Civil Engineering*, 23(1), 3-13.

Tsai, Y., Hu, Z., and Wang, Z. (2009b). "Vision-based roadway geometry computation." *Journal of Transportation Engineering*, 136(3), 223-233.

Tsai, Y., Kim, P., and Wang, Z. (2009). "Generalized Traffic Sign Detection Model for Developing a Sign Inventory." *Journal of Computing in Civil Engineering*, 23(5), 266-276.

Tsai, Y., Kim, P., and Wang, Z. (2009a). "Generalized traffic sign detection model for developing a sign inventory." *Journal of Computing in Civil Engineering*, 23(5), 266-276.

Tsai, Y., and Wu, J. (2002). "Shape- and Texture-Based 1-D Image Processing Algorithm for Real-Time Stop Sign Road Inventory Data Collection." *Journal of Intelligent Transportation Systems*, 7(3-4), 213-234.

Tsai, Y. J., and Wang, Z. (2008). "Image processing algorithms for an enhanced roadway sign data collection." *Seventh International Conference on Managing Pavement Assets*.

Tuite, K., Snavely, N., Hsaio, D. Y., Tabing, N., and Popovic, Z. (2011). "PhotoCity: Training Experts at Large-scale Image Acquisition Through a Competitive Game." *Proc., annual conference on human factors in computing systems,* Vancouver, BC, Canada.

Uslu, B., Golparvar-Fard, M., and de la Garza, J. M. "Image-based 3D reconstruction and recognition for enhanced highway condition assessment." *Proc., Proceedings of the 2011 ASCE Intl. Workshop on Computing in Civil Engineering, Miami, FL*, 67-76.

Veneziano, D., Hallmark, S. L., Souleyrette, R. R., and Mantravadi, K. "Evaluating Remotely Sensed Images for Use in Inventorying Roadway Features." *Proc., Applications of Advanced Technologies in Transportation (2002)*, ASCE, 378-385.

Viola, P., and Jones, M. "Rapid object detection using a boosted cascade of simple features." *Proc., Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, IEEE, I-511-I-518 vol. 511.

Walters, R., and Jaselskis, E. (2005). "Using Scanning Lasers for Real - Time Pavement Thickness Measurement." *Computing in Civil Engineering*, 1-11.

Wang, K. C., Hou, Z., and Gong, W. (2010). "Automated road sign inventory system based on stereo vision and tracking." *Computer-Aided Civil and Infrastructure Engineering*, 25(6), 468-477.

Wang, Y.-j., Yu, Z.-c., He, S.-b., Cheng, J.-l., and Zhang, Z.-j. "A Data-Mining-Based Study on Road Traffic Information Analysis and Decision Support." *Proc., Web Mining and Web-based Application, 2009. WMWA '09. Second Pacific-Asia Conference on*, 24-27.

Wen-Jia, K., and Chien-Chung, L. "Two-Stage Road Sign Detection and Recognition." *Proc., Multimedia and Expo, 2007 IEEE International Conference on*, 1427-1430.

Wu, C. (2007). "SiftGPU: A GPU implementation of scale invariant feature transform (SIFT)."

Wu, C., Agarwal, S., Curless, B., and Seitz, S. M. "Multicore bundle adjustment." *Proc., Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*, IEEE, 3057-3064.

Wu, J., and Tsai, Y. (2005). "Real-time Speed Limit Sign Recognition Based on Locally Adaptive Thresholding and Depth-First-Search." *Photogrammetric Engineering and Remote Sensing*, 71(4).

Wu, J., and Tsai, Y. (2006a). "Enhanced roadway geometry data collection using an effective video log image-processing algorithm." *Transportation Research Record: Journal of the Transportation Research Board*, 1972(1), 133-140.

Wu, J., and Tsai, Y. J. (2006b). "Enhanced Roadway Inventory Using a 2-D Sign Video Image Recognition Algorithm." *Computer-Aided Civil and Infrastructure Engineering*, 21(5), 369-382.

Wüller, D., and Gabele, H. "The usage of digital cameras as luminance meters." *Proc., Electronic Imaging 2007*, International Society for Optics and Photonics, 65020U-65020U-65011.

Xie, Y., Liu, L.-f., Li, C.-h., and Qu, Y.-y. "Unifying visual saliency with HOG feature learning for traffic sign detection." *Proc., Intelligent Vehicles Symposium, 2009 IEEE*, IEEE, 24-29.

Xu, Q., Su, J., and Liu, T. "A detection and recognition method for prohibition traffic signs." *Proc., Image Analysis and Signal Processing (IASP), 2010 International Conference on*, 583-586.

Xuming, H., Zemel, R. S., and Carreira-Perpindn, M. A. "Multiscale conditional random fields for image labeling." *Proc., IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR 2004)*, 695-702.

Yangxing, L., Ikenaga, T., and Goto, S. "Geometrical, Physical and Text/Symbol Analysis Based Approach of Traffic Sign Detection System." *Proc., Intelligent Vehicles Symposium, 2006 IEEE*, 238-243.

Yea-Shuan, H., and Yun-Shin, L. "Detection and recognition of speed limit signs." *Proc., Computer Symposium (ICS), 2010 International*, 107-112.

Yea-Shuan, H., Yun-Shin, L., and Fang-Hsuan, C. "A Method of Detecting and Recognizing Speed-limit Signs." *Proc., Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, 371-374.

Zamani, Z., Pourmand, M., and Saraee, M. H. "Application of data mining in traffic management: Case of city of Isfahan." *Proc., Electronic Computer Technology (ICECT), 2010 International Conference on*, IEEE, 102-106.

Zhang, C., Wang, L., and Yang, R. (2010). "Semantic segmentation of urban scenes using dense depth maps." *Computer Vision–ECCV 2010*, Springer, 708-721.

Zhang, X., and Pazner, M. (2004). "The icon imagemap technique for multivariate geospatial data visualization: approach and software system." *Cartography and Geographic Information Science*, 31(1), 29-41.

Zhou, H., Jalayer, M., Gong, J., Hu, S., and Grinter, M. (2013). "Investigation of Methods and Approaches for Collecting and Recording Highway Inventory Data."